

Physics

for the IB Diploma

SIXTH EDITION

OPTIONS

K. A. Tsokos

with additional
online material



Nature of Science

Introduction

The word ‘science’ derives from the Latin *scientia*, meaning ‘knowledge’. Historically, the term has been particularly used to describe knowledge based on clear, reproducible evidence. The implication is that the resulting knowledge is reliable and objective, and might be used to make predictions. This essentially means that whatever is claimed to be true can be proven again (and again) in similar circumstances.

The modern term ‘science’ refers to a process of creating knowledge, rather than the body of knowledge itself, but the principles underlying the process are still related to the meaning of the original Latin term. Knowledge is created through a process that must be objective, based on evidence. Whatever knowledge is generated should be true irrespective of the context, circumstances and time. However, while the aim of science is to be objective, we shall see that it is not possible to completely disconnect the process from influences of culture, economics and politics.

The process of science is based on different methods of gathering evidence, including experimentation and observation. The methodology is designed to answer specific questions or test hypotheses (testable explanations), and it may make use of models. The data obtained may lead to the construction of theories and laws. But, while science is often seen as a strictly methodological process, scientists also have to be ready for unplanned, surprising or accidental discoveries – the history of science shows that this is a common occurrence – and working as a scientist therefore requires creativity and imagination as well as structured thinking.

A universal language would facilitate and support the process of science. Use of a single language would mean that scientists worldwide could agree on what is being discussed, without misunderstandings being introduced in translation. In fact, different ‘universal languages’ are used in different areas of science. For example, many aspects of physics and chemistry are expressed in mathematical notation, chemists use chemical equations and structural formulae, and Latin is used extensively in biology and medical sciences. Nowadays, the bulk of scientific literature is in English irrespective of the native language of the scientists or the country where the research was conducted or published.

It is important to recognise that science is a dynamic process: the understandings that underlie ongoing research evolve and develop, and theories may be falsified and replaced by newer ones. The general public often do not understand this aspect of science. Many people think that science is a fixed body of knowledge and, if they see that a theory is no longer accepted or is unable to explain recent findings, they conclude that ‘science’ is unreliable. However, it is precisely the dynamic nature of science that makes it reliable and trustworthy. It demonstrates a constant striving for the best possible description and explanation, and it guarantees that the ‘current’ theory describes a phenomenon as accurately as possible given existing knowledge. The nature of science is to renew itself constantly. It is an exciting and challenging adventure where the focus lies on searching for new knowledge.

The word **scientist** was first used in 1834 by naturalist and theologian William Whewell. Before that, people who studied the natural world were known as **natural philosophers**.

Scientists may work together with technologists to create new technologies but progress in technology can be limited by current scientific theory. This may then trigger further research to solve technological quandaries. Technology and science are closely linked, but technology requires scientific understanding in order to exist and develop.

This chapter will discuss many aspects of science. It will show, above all, that science is an exciting, human endeavour with all its fallacies, weaknesses and pitfalls. The strength of science lies in the underlying process which guarantees that truth will ultimately prevail.

1 What is science?

The purpose and processes of science

The nature of science is based on a number of **axioms**, or assumptions that are seen as self-evident. These assumptions are that:

- the Universe has a reality that is independent – in other words, the Universe exists whether or not we are there to see it
- this reality can be accessed by human senses or instrumentation and understood by human reason.

The main aim of **pure science** (or basic science) is to discover what that objective reality is. This is done by collecting evidence from which conclusions can be drawn about the nature of the Universe. These conclusions may in turn lead to more questions about the Universe, meaning that new evidence needs to be gathered to answer those new questions. In summary, the nature of science is to convert the concrete (observations) into abstractions (laws and theories).

Pure science has a different aim from **applied science**, which uses scientific understanding for a specific purpose. For example, pharmaceutical scientists use their understanding of the human body and of characteristics of certain chemicals to find new medicines. Both pure and applied science can in turn contribute to the fields of technology and engineering, which focus on using and improving tools and systems to solve practical problems. The boundaries between these various fields are not distinct, and insights in one field can frequently lead to progress in another.

It is sometimes suggested that there is a single **scientific method**, but this is not correct: different methodologies are required to obtain different kinds of evidence. The type of evidence needed will depend on the question that the scientist is trying to answer, and it will influence the way in which that evidence is interpreted and conclusions are drawn. However, there must be agreement among scientists as to what constitutes a scientifically valid method. After all, what value is a finding that has meaning to only one person? Findings must be the same universally, otherwise they cannot be said to be objective and independent. For this reason, many methods are standardised, and methods must be communicated in such a way that another scientist could follow the same method to reach the same conclusions.

This section covers:

- the purpose and processes of science
- obtaining evidence
- drawing conclusions – deduction and induction
- intuition and serendipity
- scepticism
- the language of science.

Other areas of applied science include:

- electronics
- food science
- forensic science
- environmental science.



Obtaining evidence

By **evidence** we mean data about the Universe that reveal something about its nature. Evidence can be gathered using the human senses, but instrumentation and sensors are increasingly employed. Using technology to gather evidence is a more objective method and can also allow gathering of evidence not accessible to human senses. Just think of measuring temperatures in nuclear reactors or gathering data at the bottom of the ocean near black smokers. Most of the evidence gathered in fields such as astronomy would also be impossible without the help of modern technological instruments.

Evidence can be obtained using three general methods.

- **Observations:** Galileo Galilei's observations of the moons of Jupiter at the beginning of the 17th century were important evidence against the theory that everything in the Universe orbits the Earth.
- **Experimentation:** Gregor Mendel's experiments in the 19th century in which he cross-bred pea plants led to important insights into the mechanism of heredity.
- **Modelling:** modelling the current motion of the galaxies has led to the conclusion that the Universe is 13.8 billion years old.

Evidence obtained through any of these methods can be used to support a claim about the nature of the Universe, and many established scientific theories have been built on a combination of evidence obtained from all three methods.

Observations

Observation is the direct recording of data about the Universe. It is important to realise that observation does not just include seeing things with our eyes. Observation also includes using other human senses and, increasingly, instrumentation and sensors.

Our understanding of naturally occurring events is largely based on observation: think, for instance, about observations of solar eclipses or the ongoing monitoring of the extent of sea ice in the Arctic. Scientists may also observe data that can lead to conclusions about processes that have happened in the past: for example, observations of existing organisms and those found in the fossil record informed the theory of evolution. But scientists also observe events that they bring about themselves in the laboratory: for instance, holding a sample of calcium in a flame turns the flame brick-red.

It is sometimes suggested that conclusions reached through observation of naturally occurring phenomena are less valid than those reached through experimentation (see below), but this is not true. As long as the process of reasoning is sound, the conclusions reached are valid. The subsequent discovery of further evidence to support these conclusions may further strengthen the conclusions.

Some areas of science depend to a great extent on observation. An obvious example is astronomy. Observations of the radiation emitted by distant galaxies, as well as the radiation that fills the Universe, lent support to the Big Bang theory for the origin of the Universe. This well-known theory became established through the power of observation combined with structured reasoning and elaborate modelling.

The Big Bang theory

When the Big Bang theory was developed, no controlled experiments were possible to support it because of the time spans and scale of the events. However, experiments at the Large Hadron Collider at CERN (*Conseil Européen pour la Recherche Nucléaire* – the European Organization for Nuclear Research) can now recreate conditions that might resemble the early Universe, albeit for a very short time, and can be used to test theories of what the Universe was like in its infancy.

Experimentation

Observation may lead to ideas or questions that can be studied through **experimentation**. An experiment is a test designed to answer a specific question – for example, about what happens in a particular process. In an experiment, the researcher performs certain actions and observes their effects. Thus, experimentation is, in effect, a specific form of observation. In an experiment, conditions are controlled so that the researcher can be sure that the effects observed are the result of the actions carried out.

One of the key uses of experiments is to establish cause and effect – in other words, to find out whether one variable or factor has an effect on another variable or factor. The principle of this type of experiment is highly standardised in science. The researcher will make changes to one variable (e.g. the temperature) and then measure a second variable (e.g. the rate of a chemical reaction). In this way, the researcher can determine how temperature affects the rate of this particular reaction. To be sure about this result, other variables that may affect the rate of reaction must be controlled. For example, in each test, the researcher would use the same concentrations of chemicals, and follow the same procedure of mixing, stirring and so forth.

Examples of famous science experiments

In 1747, James Lind added different foods to the diet of crew members on long sea voyages. The results showed that eating citrus fruit prevented the crew from getting a disease called scurvy, which was common among 18th-century seafarers, while cider, vinegar and sea water did not prevent the disease. We now know that scurvy is caused by a lack of vitamin C.

In 1909, Ernest Rutherford designed an experiment in which α -particles were fired at thin sheets of metal. Most of the particles passed through the films, but some were deflected (changed direction). To explain these results, Rutherford suggested that the atoms in the metal consisted of a dense core at the centre – the nucleus – surrounded by an area of mostly empty space.

Forecasting the weather

A prime example of the application of models is in weather forecasting. Many factors influence weather systems, such as changes in the jet stream, sea water temperature, carbon dioxide concentrations and solar output, to name but a few. Computer models harnessing all these data have significantly improved the accuracy of weather forecasts.

Modelling

An established theory can be used to formulate a **model**. A model is any representation of an object, concept or process. There are many different types of model. Some models help us to **visualise** processes. For example, a flow chart may be used to model the pathways in human metabolism to help us see how they interact. The Bohr model of the atom is an example of a model that offers a particular way of thinking about a concept. It does not describe the atom exactly, but it describes certain features in a way that explains particular properties, such as the absorption and emission of radiation by atoms.

Modern advances in computing power have allowed the development of elaborate mathematical and computational models, and these have had an immense influence and impact. Models have become powerful tools which are capable of predicting the precise outcome of certain experiments.



Drawing conclusions – deduction and induction

Given the same set of results and background information, all scientists should come to the same or similar conclusion. Their training gives them common reasoning skills of deduction and induction.

Deductive reasoning involves using a set of general statements or observations that we know to be true to reach a logically sound conclusion. It is a ‘top-down’ process – we use general truths to arrive at a conclusion. An example of this kind of reasoning goes as follows:

Statement: *Increasing nitrate concentration in the soil causes plants to grow faster.*

Statement: *Organism X is a plant.*

Conclusion: *Organism X will grow faster if the nitrate concentration in the soil is increased.*

Thus, deduction is a reliable way of arriving at a conclusion.

However, there are many situations in which we make an observation that we cannot explain using a set of general statements we know to be true, because we do not have enough prior knowledge. In those situations, we must use inductive reasoning. This is a ‘bottom-up’ process, which involves generalising from a few specific observations to reach a more general conclusion. An example of this type of reasoning is as follows:

Observation: *All swans we have seen are white.*

Conclusion: *All swans in the world are white.*

The conclusion above **could** be true because it fits the observations. However, as it turns out, there are also black swans. This is a problem with induction: just because an event has always been observed happening in a particular way, or every known example of a particular object has certain characteristics, does not necessarily mean that this event will always happen in this way or that these objects will all have these characteristics.

Induction is therefore a less reliable method of arriving at a conclusion that is true, but it can still be a useful way of thinking. It is common to begin reasoning through induction based on a small number of observations and then to find more evidence to support the initial tentative conclusion. Charles Darwin’s **theory of evolution** is a good example of this.

In the Galapagos Islands off South America in 1835, Darwin (1809–1882) observed that finches feeding on different foods had beaks of different sizes and shapes. He then used his observations to draw conclusions about the evolution of the birds’ beaks. He thought that the environment – the availability of a variety of food sources – must have had an impact on how the beaks of these finches changed over time. Natural selection would have ensured the survival and increased breeding success of those finches best suited to a particular food type, resulting in speciation (species formation). It is clear that this type of conclusion would need further evidence, but it was the inductive reasoning based on these first observations that led to the gathering of the huge body of evidence that now supports the theory of evolution.

Einstein's intuition

Albert Einstein (1879–1955) used his intuition to work out the basic concepts of relativity. Only later was he able to develop the mathematics necessary to express his ideas and predictions. It is interesting to note that some of his predictions were proven many years later – when Einstein's ideas were published, the technology and instrumentation did not exist to test his predictions.

Intuition and serendipity

The discussion so far may give the impression that science is always a methodical business, that conclusions follow logically from evidence, and that conclusions lead in a straightforward manner to new theories and areas of research. This is by no means always the case. Great leaps forward have been made thanks to **intuition**, speculation and creativity.

Another driver of scientific discovery is serendipity, or 'happy accident'. In the pursuit of new data, scientists can come across unexpected findings in their work in the lab or in the field which can lead to great discoveries. Perhaps the most famous example of scientific serendipity is the discovery of penicillin. Sir Alexander Fleming (1881–1955) left a Petri dish containing a culture of *Staphylococcus* sp. open by mistake. The bacterial culture was contaminated by a blue-green mould, which formed a visible growth and inhibited the growth of the bacterium. The mould was isolated and purified and found to belong to the *Penicillium* genus. Somehow this fungus could make and release a substance with antibacterial activity. This substance is now known as penicillin, and it has been one of the most important life-saving discoveries in medical history.

This example demonstrates that significant scientific discoveries can be a matter of luck. However, it still takes an astute and creative scientist to recognise what she/he observes and pursue it further. As the Hungarian biochemist Albert Szent-Gyorgyi (1893–1986) said: 'Research is to see what everybody else has seen, and to think what nobody else has thought.'

Scepticism

Science is a human endeavour and therefore errors due to human fallibility and subjectivity will inevitably occur. Experimentation or observations can lead to certain claims, but scientists should initially be sceptical. Any claim should be judged only once there is good reason to believe it to be either true or false, based on solid evidence and reasoning.

Unexpected findings have been known to lead to exceptional claims. The cold fusion claim is one notable example: nuclear fusion normally requires temperatures above 10 000 000 K, so the claim that fusion could be achieved at room temperature was extraordinary. With such highly controversial claims it is easy to remain sceptical. Most scientists would not immediately accept or dismiss such findings; they would either try to repeat the experiments or wait until other scientists published similar results. Few claims are as extraordinary as the cold fusion example, but nothing that a scientist publishes should be accepted without solid evidence and reasoning that put it beyond doubt.



Cold fusion

In 1989, Stanley Pons and Martin Fleischmann reported that they had achieved 'cold fusion'. They had designed a small table-top reactor in which they electrolysed heavy water on the surface of a palladium (Pd) electrode. Their apparatus produced excess heat, which could not be explained by the chemical process that took place in the reactor. The only explanation seemed to be in line with nuclear processes and, further to this, the team reported measuring small amounts of nuclear reaction by-products, including neutrons and tritium.

Other laboratories attempted to repeat the experiments of Fleischmann and Pons but to no avail. Their findings have never been corroborated and this area of research has now largely been abandoned.

See also: http://undsci.berkeley.edu/article/cold_fusion_01

(Note that this area of research should not be confused with muon-catalysed fusion, which is an established area of research.)

The language of science

It is of paramount importance that scientists should use a common language. Science is a global enterprise and it makes sense that results can be read, understood and used by everyone around the world.

Today, the vast majority of scientific proceedings and most scientific journals are published in English, and English has also become the standard language for international conferences and congresses. This communality facilitates collaboration between scientists of different nationalities.

In addition, scientists in certain fields have developed their own terminology, notations and other conventions to make sure that they can communicate unambiguously. Medical scientists and biologists heavily rely on Latin. The physical sciences have agreements about standard units – called SI units – and notations that are used as defined by the International Bureau of Weights and Measures (BIPM). Chemistry has adopted universally understood symbols to represent the elements, and the International Union of Pure and Applied Chemistry (IUPAC) is the accepted authority on the standardisation of chemical nomenclature.

Mathematics is a powerful tool for scientists that can be considered a language in itself. Many scientific ideas in disciplines such as physics can only be expressed mathematically.

BIPM: <http://www.bipm.org>

IUPAC: <http://www.iupac.org>

Learning objectives

This section covers:

- theories and paradigm shifts
- laws
- Occam's razor
- hypotheses and falsification
- correlation and cause.

Understanding the cause of stomach ulcers

A good example of a paradigm shift is the acceptance of the cause of stomach ulcers. In the early 1980s, Barry Marshall and his co-worker Robin Warren (Nobel laureates in 2005) proved that the bacterium *Helicobacter pylori* could cause ulcers. It had previously been widely accepted that stomach ulcers could be caused by stress and other factors but not as a result of a bacterial infection. It took a long time before their theory was accepted in the scientific community.

2 Understanding of science

Theories and paradigm shifts

In science, a **theory** is defined as a comprehensive model of how a particular process or part of the Universe works. A theory may contain or be built upon definitions, facts, laws and hypotheses that have been tested.

The scientific meaning of the word 'theory' is therefore very different from the meaning it sometimes has in public understanding, referring to a vague, unsubstantiated idea. If something is referred to as a theory in science, there is no reason to doubt its validity. In fact, quite the opposite is true: established scientific theories are based on large bodies of evidence.

Examples of scientific theories

The theory of evolution describes how natural selection drives the change in inherited characteristics of living organisms over time. For example, it can predict what will happen to a bacterial population when it is subjected to the environmental presence of antibiotics. Natural selection will ensure that only those bacteria within the population that can enzymatically break down the antibiotics will survive. So the exposure to antibiotics drives the population to evolve antibiotic resistance and this feature is passed on to the offspring of those bacteria.

Isaac Newton's theory of gravitation reliably describes the gravitational pull that any two objects will have on each other, and it can be used to predict the behaviour of planets.

The atomic theory can be used to make predictions about the properties of substances on a macroscopic scale.

While individual theories concentrate on well-defined areas of knowledge, there is overlap in the facts and assumptions incorporated into different theories. This means that scientific understanding comprises a coherent body of knowledge that hangs together in a consistent way. From time to time, however, new theories emerge that have widespread implications for other theories, causing a radical change in understanding. Such a change in understanding is called a **paradigm shift**; *paradigm* is the Greek word for 'pattern'. Paradigm shifts are part of the nature and strength of science, ensuring that scientific ideas always reflect the latest evidence.

The term 'paradigm shift' was first introduced by Thomas Samuel Kuhn (1922–1996), an American physicist, historian and philosopher of science, in his book *The Structure of Scientific Revolutions*, published in 1962. Kuhn stated that scientific knowledge progressed not in a gradual way but by periodic paradigm shifts. He described these shifts as 'universally recognised scientific achievements that, for a time, provide model problems and solutions for a community of researchers'. Thus, a paradigm shift represents a major move away from a previously held notion. It provides the scientific community with novel views, approaches and explanations which, up to that time, had been absent or in some cases might have been considered heresy.



Paradigm shifts do not necessarily make ‘old’ theories invalid. At the beginning of the 20th century, Albert Einstein’s theory of relativity represented a paradigm shift relative to Newtonian mechanics. However, Newtonian mechanics are still perfectly applicable in many situations. Here, the new paradigm offers a deeper and wider understanding, but it does not make the old paradigm obsolete.

Laws

In science, a **law** is a statement that describes a particular behaviour. Laws are derived from repeated observations or experiments and often describe a relationship between two or more variables. A law states that the same result or phenomenon is always observed under the same conditions, and it can therefore be used to make predictions. Because laws need to be universal and should be easily understood across languages and cultures, they are often expressed as mathematical formulae or equations.

Note that, in contrast to a theory, a law does not **explain** a phenomenon, it only **describes** one.

Examples of scientific laws

- **Newton’s second law** states that the acceleration of a body is proportional to the net force on the body and inversely proportional to the mass of the body. This is expressed mathematically as $F = ma$.
- **The law of conservation of mass** states that the amount of matter does not change in a chemical reaction – there will be the same amount of matter after the reaction as there was before it.

Occam’s razor

In formulating a law or theory, a scientist should strive for the simplest form that fits the available evidence. This essentially means that the simplest explanation for a phenomenon is assumed until further evidence suggests that a more complicated explanation is needed. This principle is known as **Occam’s razor**, attributed to William of Occam (c.1287–1347), a philosopher and theologian.

As an example, let us imagine that a scientist has conducted an experiment on the effect of nitrate concentration on the growth of plants. The results show that nitrates cause plants to grow faster. The scientist could now start to formulate a theory about the relationship between the growth of plants and nitrates. The simplest explanation is that the nitrates are absorbed by the plants and used in a way that is beneficial to their growth. A more complicated explanation might be that the nitrates are poisonous to worms, that the lack of worms causes the population of moles to decline, and that the absence of moles means that there is less damage to plant roots, allowing the plants to grow better. Both of these explanations fit the observation, but the simpler explanation should initially be used to guide further investigations into the exact effect of nitrates on plants.

Falsifiability

Sir Karl Popper (1902–1994), an Austro-British philosopher of science, alleged that falsifiability marks the boundary between science and pseudoscience (see Section 5). He argued that research findings based on an unfalsifiable hypothesis could only be considered pseudoscientific and thus could not be taken as reliable.

Hypothesis or law?

A hypothesis may look similar to a law, but it is important to recognise that the two are very different. A hypothesis forms the basis of an investigation and can be rejected on the basis of a single experiment. A law has been established through repeated observations or experiments, and it forms part of accepted scientific understanding.

Hypotheses and falsification

Scientific knowledge develops through the testing of hypotheses. A **hypothesis** is a testable statement or prediction. A scientist may formulate a hypothesis based on an idea they have about how the world works, for example based on particular observations or prior experiments. The hypothesis is then tested through experimentation.

Hypotheses should be formulated so that they are **falsifiable**. This means that the hypothesis must be phrased in such a way that an experiment can be designed to prove it wrong. The following example will illustrate this.

Suppose that a team of scientists has observed that, in a paddock, plants located closer to an area where cows frequently urinate grow larger than those in other areas. Urine contains urea, a nitrogen-based compound which can be converted to nitrates by nitrifying bacteria. Based on these observations, the scientists might propose that nitrate positively influences the growth rate of plants. They might propose the following hypothesis:

Increasing nitrate concentrations in the soil increases the growth rate of plants.

Note that the hypothesis is a statement, not a question.

Is this hypothesis testable? Yes it is, because we can design an experiment in which we vary the nitrate concentration and measure the effect on plant growth rate. Is the hypothesis falsifiable? Yes, because if plants do **not** grow faster in soil with higher nitrate concentration, the hypothesis will be proven wrong.

Note also that it is not possible to prove that a hypothesis is true, only that it is false. Here, we can show that increasing nitrate concentrations in the soil increases the growth rate of plants **in this particular study**, but that does not guarantee that it is always true. This is the induction problem (see Section 1). The hypothesis can, however, be supported if the same effect is observed over and over again, and if a plausible mechanism is found for how nitrates might stimulate plant growth.

The scientific process can be summarised as follows.

- A scientist will use inductive and deductive reasoning based on observations and/or experimentation to arrive at certain conclusions.
- He or she may then begin to formulate a theory.
- This theory needs to be tested; the scientist proceeds by formulating hypotheses.
- Experimentation based on these hypotheses will yield further observations and conclusions.
- The evidence found by the scientist will give rise to further (testable) questions, which will lead to further experimentation.

In real laboratory life, this process may take many years and involve many people, since each of the steps sometimes requires difficult and lengthy experiments. However long it takes, it brings us full circle and demonstrates how science progresses in a dynamic and developmental way.

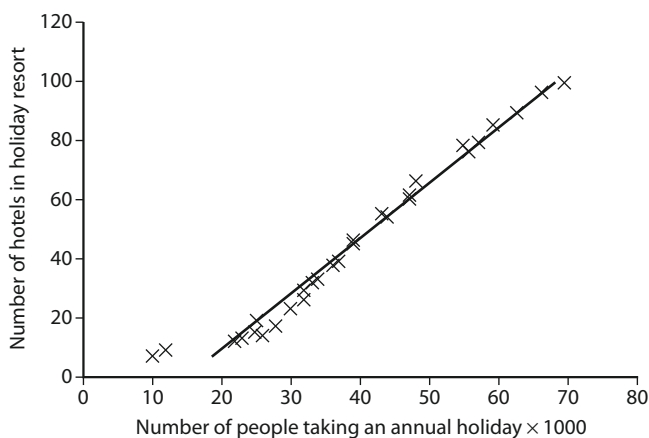


Figure 1 This graph shows a positive correlation. If more people take a holiday in a resort, then that resort will have more hotels.

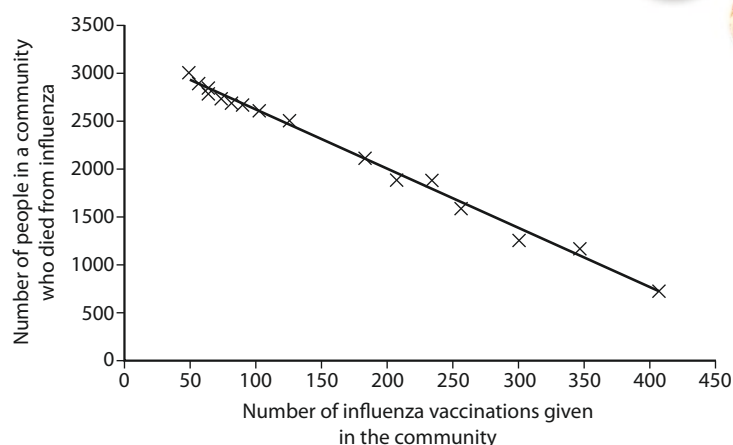


Figure 2 This graph shows a negative correlation. In communities where more people were vaccinated against influenza, fewer people died from the disease.

Correlation and cause

Correlation is a reliable statistical link between two variables. This means that, as one variable increases, the other variable either also increases (positive correlation, Figure 1) or decreases (negative correlation, Figure 2).

It is important to realise that, while a strong correlation between two variables may suggest that there is a causal relationship (i.e. that a change in the first variable directly causes a change in the second), this is not necessarily the case. For example, look at the graph in Figure 1. It could be that increases in the number of holidaymakers in certain resorts caused more hotels to be built there. However, the reverse could also be true: holidaymakers could be attracted to certain resorts because they have a lot of hotels.

Observation of a correlation therefore often warrants further studies to establish **causation**. The best way to establish causation is through a carefully controlled experiment in which the effect of altering one variable is measured, but this is not always possible. For example, if a study has shown a positive correlation between eating fried food and getting bowel cancer, it would be unethical to feed different groups of people different amounts of fried food and see how many develop bowel cancer. So, other methods are sometimes needed.

One way of supporting the case for a causal relationship is to propose a plausible mechanism by which one variable could have an effect on another. For the fried food/bowel cancer example, a plausible mechanism might be that certain chemical compounds in fried food cause DNA mutations in cells in the bowel. These mutations may involve genes that suppress cancer. These events, combined with a weakened immune system and other genetic factors, may lead to bowel cancer. This is a complex picture, but it is possible to carry out experiments to establish parts of the mechanism. For example, chemicals from fried foods can be added to human bowel cells *in vitro*, to see if the cells turn cancerous.

Other methods used to obtain evidence for causal relationships in medical studies include sampling, cohort studies, case control studies, double-blind tests and clinical trials. All of these are basically surveys with large numbers of people (patients) with similar backgrounds, diets, age and so on, so that as many variables as possible are controlled. Statistics (see Section 3) are an indispensable tool for the analysis of these data.

The term *in vitro* is Latin for ‘in glass’. It is used to describe studies carried out on parts of organisms outside the living organism. Studies on living organisms are termed *in vivo*, which means ‘within the living’.

This section covers:

- qualitative and quantitative data
- repetition and replication
- errors
- statistics
- cognitive bias
- outliers
- databases.

3 Objectivity of science

Qualitative and quantitative data

Data, or evidence, can be in two basic forms: quantitative and qualitative. **Quantitative data** are based on measurable quantities and are therefore numerical. They are measured using tools or instrumentation yielding values with (standardised) units. For example, the temperature of a reaction mixture (in °C) or the volume of gas produced in a chemical reaction (in cm³) constitutes quantitative data.

Qualitative data deal with apparent or implicit qualities and are expressed in words. They are usually observations, made either in an experiment or from an examination of something. The following are examples of qualitative data: 'the reaction mixture turned cloudy'; 'when the two objects collided, a loud noise was heard'; 'this type of insect lifts its wings when threatened'.

Quantitative data are usually objective and more suitable than qualitative data for accurately describing phenomena and making predictions. Because they are numerical, quantitative data can be mathematically analysed to establish links between variables and to identify patterns. On the other hand, qualitative data are seen as more subjective, and the research for this type of data gathering is far more difficult to repeat or confirm. This is not to imply that qualitative research is not valid, but it is less likely to yield theories or laws that are applicable to all humanity or valid throughout the whole Universe. That is why scientists prefer to rely on quantitative data.

Repetition and replication

Science and data are inextricably linked. Data need to be reliable so that realistic and trustworthy predictions can be made, and the reliability of data can be improved by making repeated measurements.

It is therefore good practice for scientists to take repeated measurements, by performing the same experiment multiple times. If an experiment is **reproducible** (i.e. it gives the same result each time it is repeated), then we can have confidence in the results. If the values measured in each experiment are close together, then we say the measurements have high **precision** (see below).

In addition to scientists repeating their own results, it is also important that results are **replicated** by other scientists in different settings. If the results cannot be replicated, this might mean that there was an error inherent in the original procedure (see below), leading to false results. Replication is important to show that results are **accurate**, that they are a true reflection of reality. Figure 3 illustrates the difference between precision and accuracy.

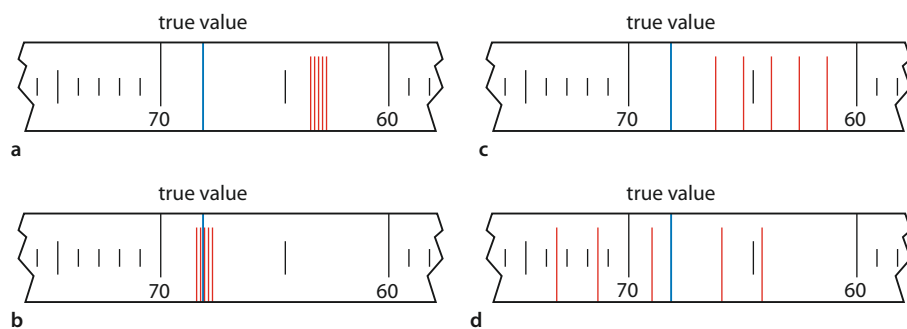


Figure 3 The difference between precision and accuracy. Results are shown in red and the true value measurement is indicated in each diagram by a blue line. The results in **a** have high precision (they are close together) but low accuracy (they are not close to the true value of the measurement). The results in **b** have high precision and high accuracy. The results in **c** have low precision and low accuracy. The results in **d** have low precision and high accuracy (because the mean value is close to the true value). Despite the high accuracy, this last set of results would still be considered poor data.

Some large experimental set-ups, such as the Large Hadron Collider (LHC) in Geneva, Switzerland, generate vast amounts of data which are difficult to interpret. In order to increase the reliability of these data, it makes sense to replicate the experiments, but that is a costly affair. The LHC therefore has replication built in: it has multiple detectors. The detectors perform the same experiments but they are run by different groups so, when they produce the same result (such as identifying the Higgs boson), we know that the result is reliable.

Errors

However carefully an investigation is carried out, it is always possible for errors to occur. Scientists must have an in-depth understanding of how and why errors occur and must consider to what extent any errors may have affected the data. Errors mean that quantitative data are not always as objective and accurate as one might think, but careful experimental design can reduce the number and impact of errors. For example, temperatures read off an analogue mercury thermometer may vary slightly depending on who takes the measurement. However, if a digital thermometer is used, the readings should all be the same. Scientists therefore rely heavily on equipment to record data. Any measuring or recording equipment used needs careful calibration to ensure that the readings are accurate and standardised, but built-in errors may still exist.

There are two main types of error: random and systematic.

- **Random errors** are caused by variables that cannot be controlled and by limitations in the measuring apparatus. For instance, if you are using a balance to measure a mass of sodium chloride, random errors in the measurement might be caused by the movement of air in the room or by friction between the mechanical parts of the balance.
- **A systematic error** is a bias in measurement that is inherent in a procedure or measurement. For example, you might measure the mass of sodium chloride using a balance that has not recently been calibrated and that consistently records a mass that is 1.00 g too high.

The two types of error affect measurements in different ways. Random errors will affect each measurement differently. The value recorded may be higher or lower than the actual value, and the difference from the actual value may be large or small. The repeated measurements will be randomly

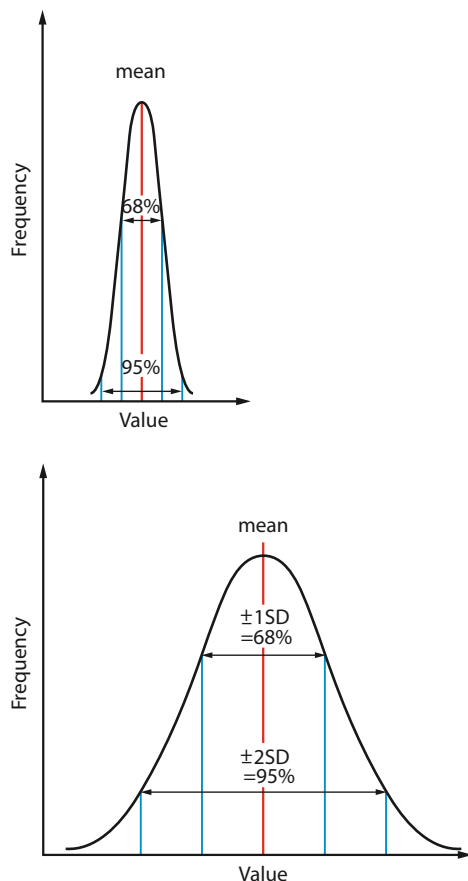


Figure 4 Two different normal distribution curves. A normal distribution is tall and narrow if the data values are close together, and flat and wide if the values are more spread out. In all cases, 68% of values fall within 1 standard deviation (± 1 SD) of the mean, and 95% of the values fall within 2 standard deviations.

Chi-squared test

The chi-squared test is used to evaluate the outcomes of genetic cross-breeding experiments. Scientists look at the appearance of certain characteristics in successive generations of an organism. They then compare those results with the results they would expect to see if the characteristics had a particular genetic basis. This can rule out or support hypotheses about the genetic basis for characteristics.

distributed around the actual value. The result is that random errors affect **precision**, or how close together repeated measurements are (see Figure 3). This means that the presence of random errors is quite easy to spot – they cause a spread in values for repeated measurements.

Conversely, systematic errors always affect measurements in the same way. If an instrument is calibrated incorrectly, it will consistently give measurements that are the same amount higher or lower than the actual value. Therefore systematic errors affect the **accuracy** of the measurements, or how close they are to the true value (see Figure 3). Systematic errors are fairly easy to spot if a literature value exists for a particular measurement but they can be much harder to spot if there is no accepted value. This is one of the reasons that it is so important that studies are replicated by other groups.

Statistics

Since errors are impossible to avoid, scientists rely on **statistics** to get a better understanding of what a ‘normal’ range is and which values are to be considered ‘outliers’ or false readings. Statistics is a branch of mathematics that concerns itself with the collection, presentation and interpretation of data, and statisticians have developed tools that help scientists to predict and assess the validity of comparing sets of data.

Statistics facilitate the summarising of large sets of data. Among other things, statistics make use of three forms of average – the mean, median and mode – each of which conveys a different aspect of the data set. The **mean** is a mathematical value obtained by dividing the sum of a set of values by the number of values in the set. The **mode** is the most frequently occurring number. The **median** is the middle value when all values are ranged in order. Different situations may require the use of different averages.

It is often found that data form a **normal distribution** (Figure 4). In a normal distribution, the measured values are distributed evenly around a central, most probable, value and the mean, mode and median values are all the same. The **standard deviation** (SD) is an indication of the spread of the data around the mean value in a normal distribution (Figure 4). If a series of repeat measurements has a high SD, this means that there is a wide spread in the data, indicating that the measurements have low precision.

There are many statistical tests that help scientists to establish whether correlations exist between variables. We can use the **chi-squared test** to compare observed data with data that we would expect to see if a certain hypothesis were true. If there is a significant difference, this proves the hypothesis false. A **t-test** is widely used to assess whether two sets of data are statistically different from each other, based on the means and standard deviations of the two data sets. Imagine, for example, that a road tyre company wants to know if their new tyre gives shorter stopping distances under braking. They will make repeated measurements (in controlled conditions) of stopping distances using the new and the old tyre. A **t-test** will show whether there is a statistical difference in stopping distance between the two tyres.



Levels of confidence

A **level of confidence** is an indication of how sure the scientist is that a true value lies in a particular interval. For example, a report might state that the concentration of arsenic in a sample of drinking water is $0.072\text{--}0.081\text{ mg dm}^{-3}$ with a level of confidence of 95%. This means that, according to the statistical calculations, we can be 95% sure that the true value of the concentration lies within that range or **confidence interval**.

The confidence interval can be calculated for any level of confidence, although 95% is common. The range of the confidence interval is an indication of the precision of the measurement. Repeating the experiment can reduce the range of the confidence interval.

Error bars and best-fit lines

When scientists depict data in graph form, error bars and best-fit lines are often displayed as well. An **error bar** is a vertical line drawn through a data point and it indicates the variability for that point. It can display the range (the minimum and maximum values measured), the standard deviation or the confidence interval for a particular confidence level. This allows other scientists to assess objectively if the data presented indeed give rise to the conclusions. For example, if the error bars between two data points do not overlap, this is a good indication that they are significantly different. Figure 5 shows what error bars look like. In this case, we can be more confident about the accuracy of the third data point than that of the first or second, because the error bar is shorter.

Best-fit lines are used to make the interpretation of a graph easier. They are widely used in scatter graphs where two variables are plotted against each other. The patterns that arise are often difficult to interpret so a best-fit line can highlight a trend. Figure 6 shows two examples of best-fit lines; note that the best-fit line is not necessarily a straight line – the form of the line depends on the relationship between the variables.

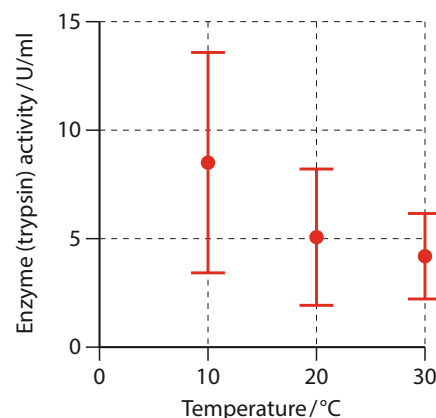


Figure 5 Error bars give an idea of the variability of the measurements obtained. In this graph of trypsin activity versus temperature, the error bar for the third data point is the shortest, which means that this value is likely to be more accurate than the other two.

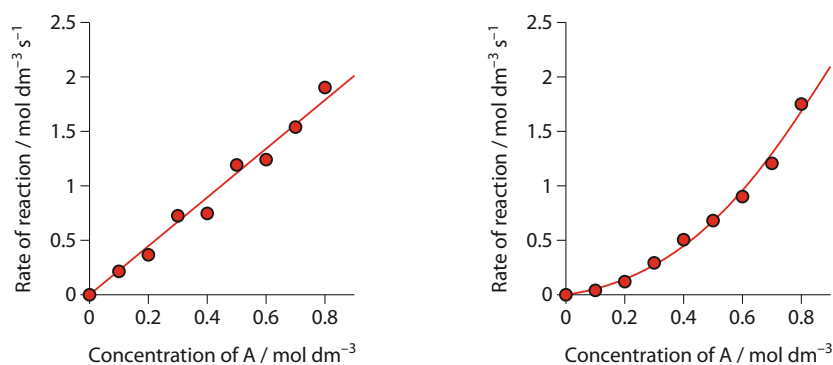


Figure 6 Best-fit lines on graphs measuring rates of reaction.

A best-fit line should be drawn so that the total distance between the data points and the line is as small as possible. A best-fit line allows other scientists to assess the data objectively and adds to the reliability of the data.

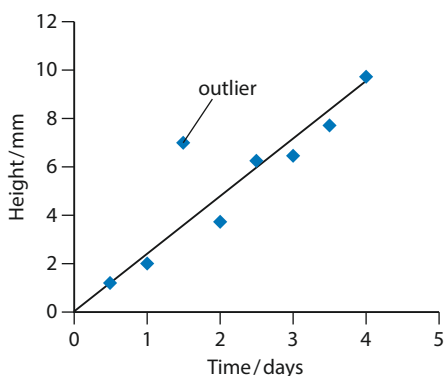


Figure 7 Drawing the best-fit line through this graph showing the growth of alfalfa seedlings makes it obvious that one of the data points is an outlier.

Examples of scientific databases

There are a large number of DNA sequence databases, many of which can be searched through the **National Center for Biotechnology Information (NCBI)** website: <http://www.ncbi.nlm.nih.gov/nuccore>

Online Mendelian Inheritance in Man (OMIM) is an online database with information on the majority of inherited diseases: <http://www.ncbi.nlm.nih.gov/omim>

The **CERN** website provides updates in the field of nuclear physics: <http://home.web.cern.ch/>

Chemical Abstracts Service (CAS), a division of the American Chemical Society, is the world's authority for chemical information: <http://www.cas.org/>

Cognitive bias

Cognitive biases are ways in which we tend to make errors of judgement in different situations. Scientists need to be aware of biases that might affect how they interpret results, so that they do not come to the wrong conclusions.

An important bias to recognise is **confirmation bias**. This is the tendency to dismiss or disagree with information that does not fit with our understanding or theories, and to favour information that agrees with what we already thought.

Imagine, for example, that a scientist obtains a set of results that confirm his favourite theory, but he hears about a different group which has obtained results that deny the theory. He may try to find errors in the other group's results to show why they are wrong but may not consider whether similar errors might exist in his own data. Alternatively, another scientist may have produced an unexpected result that does not fit with anything she has seen before. She may easily dismiss it as an error, but it is possible that this is an important new finding. These are instances of confirmation bias.

Outliers

When taking measurements, it is common for some findings to seem to be well outside the normal range. Scientists call these data **outliers**. These are often caused by random errors but, in some cases, such results are true findings, indicating that there is a larger range than expected.

Figure 7 shows an outlier in a set of results. The graph seems to confirm that most data points lie in the range of 0 to 4 (x -axis points) with values of between 0 and 10 (values on the y -axis). The data point at 1.5 on the x -axis is much further from the best-fit line than the rest of the data points, indicating that it does not fit with the expected model or theory. It is understandable, when encountering such a finding, to dismiss it as an outlier.

But should we always discard outliers? In nature, exceptions are not uncommon, therefore outliers and unexpected findings are not particularly unusual. Sometimes they can lead to new discoveries, theories and models, so scientists should remember the existence of confirmation bias and pay special attention to these 'flukes'. They must maintain a balance between healthy scepticism and too readily accepting their own favoured theories. In the example in Figure 7, the correct course of action would be to repeat the experiment, to find out if the outlier is a true result.

Databases

In some areas of science – for example, meteorology and particle physics – scientists have to analyse thousands or even millions of sets of data. The huge increase in computing power over recent decades has allowed this. In more and more areas of science, data are being stored in vast databases, and computer programs are used to analyse the data to find trends, patterns, similarities, correlations or causal relationships.

4 The human face of science

Collaboration and community

Science in the 21st century is very much a collaborative and global affair. The complexity of the problems we face – such as how to develop sustainable fusion energy, curing cancer, dealing with the greenhouse effect, the energy crisis – all require a collaborative and transdisciplinary approach. This approach can be successful because of the consistency in the training of scientists. A biochemist from Zambia would do her research in the same way as one who received his training in Australia.

Collaboration extends to scientists working with engineers and technologists. The complex problems mentioned above contain pure research questions as well as applied aspects. For the latter, technologists and engineers are needed to translate the pure research findings into practical applications. For example, a team trying to find a cure for melanoma (a type of skin cancer) may consist of biochemists, medical doctors, radiologists, molecular biologists, chemists, physicists, pharmacists and technicians.

Collaborative work can involve laboratories from a number of different universities, and from industry, hospitals and other institutions. This transdisciplinary, and often international, community enables teams to mount a concerted effort to tackle a problem from many angles and using a variety of approaches. It brings together people from different backgrounds and with different skill sets but with a common goal. This helps to ensure that the research programme as a whole is open-minded and unbiased: any prejudices that might exist within an individual scientist or team are counterbalanced by the presence of people with different points of view.

This approach increases the chances that a solution to a particular problem will be found. It is extremely rare these days that an individual scientist is capable of solving very complex problems. A collaborative approach also allows more efficient use of equipment: not all labs involved in a project need to purchase the same expensive tools. However, even though the sharing of equipment may bring down the overall costs, science remains extremely costly. Very large international projects are only possible because many nations collaborate and contribute funding.

Examples of international collaboration

In 2003, the first complete human genome was published by the **Human Genome Project (HGP)**. This was the result of 13 years of work coordinated by the US Department of Energy (<http://energy.gov/>) and the National Institutes of Health (<http://www.nih.gov/>). Other contributions to the project came from the Wellcome Trust (<http://www.wellcome.ac.uk/>), as well as groups in Japan, France, Germany and China. You can access the DNA sequence online: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml

This section covers:

- collaboration and community
- how scientists publish their work
- intellectual property
- science, ethics and the precautionary principle
- honesty in science: plagiarism and other forms of cheating
- funding and political influence.

The **Intergovernmental Panel on Climate Change (IPCC)** brings together more than 2500 scientists from around the world. It was established in 1988 to provide a clear scientific world view on the current state of knowledge about climate change and its potential environmental and socio-economic impacts.
<http://www.ipcc.ch>

The **Large Hadron Collider (LHC)** is the biggest man-made experiment in the world, housed at the CERN (the European Organization for Nuclear Research) laboratory in Geneva. Scientists of more than 100 nationalities based all over the world work together to interpret the results obtained from the LHC.
<http://home.web.cern.ch/>

An overview of scientific journals can be found here: <http://www.sciencedirect.com/science/journals>

Note that this list is by no means exhaustive.

How scientists publish their work

Scientists publish their results in scientific journals to make them available for other scientists to read and use. There are thousands of journals worldwide, both hard copy and online, where those findings may be published.

When scientists intend to publish in a scientific journal, the paper is subjected to **peer review**. This means that the paper is read and criticised anonymously by fellow scientists (peers). They will assess and check several aspects of the paper: whether the findings are novel enough for publication, whether the correct procedures have been followed, that there is no indication of plagiarism and that the report is properly referenced. In addition, they might look for conflict of interest or whether similar results have been published before. An example of a conflict of interest is when a paper demonstrates the efficiency of a new drug and the funding for the project comes from the company who produced that drug. In such cases, the results should be carefully checked for any sign that they are presented dishonestly. If all the required conditions are met, the reviewers will give the journal's editor the go-ahead for publication. Once published, the paper may be quoted by other researchers in the same field. If the findings are exceptional, they may also be quoted in the national press or on online news sites.

With the advent of online publications, an increasing number of journals are being made available for free, and a new form of peer review has emerged. Traditional journals use a team of in-house editors and trusted outside scientists to review a paper before it is accepted for publication by the journal. In the case of online publications, such as those published by the Public Library of Science (PLOS), all scientists (peers) are free to review and comment on the data. This is an open, transparent procedure rather than the closed approach used by journals. Publications are a quantifiable indication of the productivity of science.

Besides publishing in journals, scientists also present their findings at (international) conferences. These presentations usually communicate initial findings that have not yet been included in a full paper and are generally not peer reviewed, although some conferences will assess the quality of the presentations before accepting them.

PLOS

The Public Library of Science (PLOS) is a non-profit publisher [...] with a mission to accelerate progress in science and medicine by leading a transformation in research communication. (Source: <http://www.plos.org/>)

PLOS has gained quick recognition and is now a sought-after publication platform. It is freely available online.



Intellectual property

Scientists are employed by universities, institutes, hospitals and other organisations where they work in teams. Their employers often demand that they sign an agreement as part of their employment contract which gives all the intellectual rights to their discoveries to the organisation they work for. If a research finding has an applied aspect – for instance, if a new drug could be developed – it can be lucrative to apply for a **patent**.

Applying for a patent is expensive, but it is financially attractive because a patent grants the exclusive rights to use or sell the new discovery to a person or company for a period of up to 20 years. Often, all the monetary rewards go back to the company. In the case of some drugs, the financial gains for the (pharmaceutical) company can be enormous. It is estimated, for example, that the drug Tagamet (used in the treatment of stomach ulcers) earned SmithKline Beckman Corporation approximately US\$1 billion per year in the 1980s. The possibility of patenting a discovery helps make it attractive to companies to invest huge sums of money in developing new drugs.

A considerable downside to this practice is that a patent gives a pharmaceutical company a monopoly position for a particular type of drug. Within certain limits, the company can charge whatever it chooses. Of course, a company needs to recoup its costs – to get medical drugs approved for the market is a lengthy and expensive process – but, as a result, the newest drugs are expensive. This means that many people, such as those living in developing countries, are denied access to the latest medication.

Science, ethics and the precautionary principle

The field of **ethics** (or moral philosophy) deals with what is right and what is wrong. You may wonder how science can be right or wrong when it only tries to get to the truth. But to discover some truths it may be necessary to do things we consider morally wrong, which is not acceptable. For example, nobody would condone the use of human babies in medical experiments.

Scientists therefore have to be aware of the ethical implications of their work. In the first instance, they must consider the ethical aspects of their research design. To this end, governments and institutions have strict guidelines for scientific experiments involving humans or animals, and such research has to be approved by ethics committees.

But scientists must also consider the ethical aspects of the ways in which their work can be applied. Discussions on such subjects affect the wider public and they are frequently carried out in political and public forums. An example of an issue that raises widespread ethical questions is gene therapy. This technology involves changing a person's genetic make-up and has been accepted as a means of curing certain genetic diseases. However, one might ask whether it is acceptable to change a person's genes. And, if it is in some instances – for medical benefit – where should the line be drawn? Would it be acceptable to use gene therapy to 'improve' someone's behaviour or appearance, for example?

Science deals with all aspects of human life and it has the potential to solve many pressing problems. It is undeniable that science, with its partner technology, has been essential in bringing progress to many

Many thousands of patents are awarded worldwide every year. The European Patent Office (<http://www.epo.org/index.html>) has a searchable database.

areas, and this has led some scientists to proclaim that everything can be (re)solved by science. Many developments have undoubtedly been to humanity's advantage, but some inventions have the potential to be used for harmful purposes, for instance in warfare. The discovery and technological development of nuclear fission is a well-known example: it has led to the development of important nuclear power plants, but it has also led to nuclear weapons.

Scientists must therefore consider carefully whether their research could have long-ranging and far-reaching effects, in which case an in-depth discussion should precede further work in this area. Such discussions should involve policymakers as well as scientific experts, and they should include risk assessments and plans for how to manage the risks identified. If there is good reason to believe that a new technology may have harmful effects but the evidence is not clear, then the **precautionary principle** may be applied. This ensures that measures will be taken to protect the public from any risk until new evidence shows that the risk is of an acceptable level. For example, if there is good reason to believe that a particular pesticide is responsible for a reduction in bee populations – which is a very serious problem because bees pollinate many important crops – the precautionary principle states that that pesticide should be banned until the manufacturer can prove that it is not in fact harmful.

Honesty in science: plagiarism and other forms of cheating

Society expects those involved in searching for the truth to have integrity and honesty when it comes to publishing their results. There is an expectation that the data should be honestly represented, not manipulated to better fit the theory, and that any findings used to corroborate or support these data which are not the scientist's own should be properly referenced. There is considerable pressure on scientists to publish – it increases their chances of being promoted, getting more funding, and so forth – and, unfortunately, this pressure can lead to cheating.

Manipulation of data

Doctored data to make them better fit the theory or support a hypothesis is unfortunately not an uncommon occurrence. One well-documented example concerned Marc Hauser, an evolutionary biologist and professor at Harvard University. After students working in his lab reported that data in his papers were falsified, the university investigated Hauser's possible scientific misconduct. An external investigation confirmed the allegations and Hauser ultimately resigned from his post in 2011.

Falsifying data is totally unacceptable and it can have wide-ranging implications for the work of other scientists. Scientists' work is frequently based on earlier findings, and time and money are invested to repeat those findings or to corroborate them. If it is then revealed that the research was based on fraudulent data, a lot of work has been wasted. There may also be a risk of direct harm, for example if doctored data in medical research were to lead to an unsafe drug being tested on humans.



Plagiarism

Properly quoting other people and referencing work that is not the scientist's own are the norm in scientific publications. Not complying with this convention is a form of **plagiarism**. The advent of specialist computer software has made it relatively simple to assess if a text has been plagiarised. Any form of plagiarism is taken very seriously and there have been some widely publicised cases of plagiarism leading to the dismissal of the perpetrators.

Funding and political influence

Pure research is mostly funded by public institutions and governments. Research grants are available and scientists wishing to work on a particular topic have to submit research proposals which are vetted by peer review. It is a highly regulated and standardised process. Funding is limited and decisions regarding which proposals receive funding may be influenced by political considerations. Scientists therefore have to be able to make a strong case for why their research proposal is important.

Not all scientific research is conducted in publicly funded institutes, though. The defence industry and the pharmaceutical industry employ thousands of scientists who work in closed, protected conditions. The research conducted here is mostly applied research, with a fixed goal in mind. Although working for these organisations may have advantages, there are usually certain conditions imposed. Scientists are limited in what they can discuss, and publishing their findings is restricted or forbidden. The intellectual property rights or patents coming out of this research remain with the company or the defence department.

Advances in science and technology can have significant economic and political implications for a nation. For example, if research into the use of nuclear fusion energy shows that this type of energy is feasible in the very near future, it may impact on the levels of employment in oil and other energy-related industries. Politicians may not wish to see these findings published. They may have a number of valid reasons for this but it demonstrates that science can be influenced by politics.

The Lysenko affair (see right) is a good historical example of how politics may influence scientific ideas. The debate around climate change is a current, highly politicised issue. Scientists themselves find it hard to reach a consensus on how climate change will develop, how it will affect us and at what rate. The debate in science is based on the interpretation and extrapolation of data and not on personal feelings. Decisions by politicians may be influenced by the length of their terms in office or the strength of a lobbying group. However, a policy to deal with the effects of climate change, by the very nature of the problem, must be long term and demands international collaboration. Such policy decisions should ideally be based on science alone, though in practice this is hard to achieve.

The Lysenko affair

Trofim Lysenko (1898–1976) was a Soviet biologist and agronomist of Ukrainian origin who rejected Mendelian genetics. He believed that characteristics acquired by an organism during the course of its life would be passed on to the next generation and he made suggestions for improving the growing of crops based on this theory.

In the 1930s and 1940s, Joseph Stalin's forced collectivisation of the agricultural sector in the Soviet Union (USSR) caused massive production loss and resulted in famine. The country could no longer feed its own population. Lysenko's research into crop improvements was supported by the Soviet leadership and earned him the post of Director of the Institute of Genetics within the USSR's Academy of Sciences. This position allowed him to exercise political influence and power to entrench his anti-Mendelian doctrines further in Soviet science and education. Ultimately, Lysenko's theories were outlawed in 1948.

This section covers:

- science and the public
- fallacies
- pseudoscience.

Examples of scientific topics that have been the subject of public debate include:

- genetically modified foods
- nuclear energy
- climate change.

5 Public understanding of science

Science and the public

Science is inextricably linked with our lives. Communication, transport, the internet, what we eat, medicine – each and every aspect of our lives is influenced by science. It is therefore helpful if members of the public have a basic understanding of the nature of science. With that, they can make informed decisions for themselves and contribute constructively to public debate on matters related to science.

You may like to research one of the topics listed here, to find out how non-scientists have contributed to the debate.

- What kind of understanding do you think non-scientists need in order to develop an informed opinion on scientific topics?
- How have these debates been shaped by people who may not have a good understanding of the issues?

Communication between scientists and the public may be complicated by the use of scientific terminology as well as by different interpretations of certain terms. For instance, as we have seen, the public and scientific understanding of the word ‘theory’ are different: contrary to the general understanding of it being just something that might be possible, in science it denotes a model or set of laws which can be used to make predictions. The theory of fluid dynamics, developed by Swiss mathematician Daniel Bernoulli (1700–1782), plays an essential role in the development of aeroplane wings. If it were just a ‘theory’ in the lay public’s understanding of the word, it is highly unlikely that many people would ever board an aeroplane. So, explaining the terminology and its context is a good starting point for clear communication.

Another area that often causes confusion is statistics, mathematics and risk. This is partly due to some people not having a good understanding of these subjects and partly because the media frequently present results in a dramatic and sensational way, to grab readers’ attention. For example, a headline could state that drinking fizzy drinks increases the risk of pancreatic cancer by 90%. This sounds extremely serious. However, in practice, it may mean that the lifetime risk of developing pancreatic cancer increases from, say, 1.5% to 2.9%. This would still be significant, but it does not mean that drinking fizzy drinks will almost definitely give you cancer, as some people may think from reading the headline. Moreover, this result may be based on a single study that has not yet been replicated. It is therefore vital that scientists do what they can to present results in such a way that the public can understand the real risks involved. It is also important that the public have a good enough understanding of these subjects to be able to interpret the real implications of reported studies.

The idea of knowing something for certain can also be problematic. Statistics can give us a certain level of confidence in our results but it is rarely possible to know something with 100% certainty. In fields such as biology and medicine, there are many exceptions to rules. A scientist would therefore be lying if they stated that something will always be the case, that the outcome is a dead certainty. However, the public often want a definite answer.



Fallacies

A **fallacy** is a misleading or false argument, something that does not logically support the case a person is making. Both scientists and the public need to be aware of common fallacies, so that they can recognise them in scientific debate and avoid using them themselves. Common causes of these flaws in logic are an ignorance of scientific methodology and unstructured critical thinking.

The following are the most common forms of fallacy.

- **Confirmation bias** has already been introduced. In public debates about science, confirmation bias can affect whether members of the public accept a scientific message. For example, someone who believes in acupuncture is unlikely to accept a report that states that acupuncture does not have any medicinal effect.
- **Hasty generalisations** occur when people base a broad conclusion or theory on just a few observations. This is essentially an extreme form of inductive reasoning. A particularly harmful hasty generalisation in the public understanding of science is to conclude that all science must be untrustworthy because some theories in the past have turned out not to be true.
- Related to hasty generalisation is the use of **anecdotal evidence** based on subjective ‘evidence’ or ‘hearsay’. An example would be along the lines of ‘My father smoked until he was 90 and he never had lung cancer’ as an argument against the established causal link between smoking and lung cancer.
- The **post hoc ergo propter hoc fallacy** assumes that, if an occurrence A is seemingly directly followed by an event B, B must be caused by A. An example might be if a mobile phone mast was erected in an area and a few people in that area subsequently get cancer. The fallacy may lead people to conclude that the mast caused the cancer, but it is possible that it is just a coincidence. This is a very common fallacy and relates directly to the correlation versus causation debate.
- In the **straw man fallacy**, side A in a debate distorts or misrepresents the argument put forward by side B and attacks the distorted argument rather than side B’s actual argument. By doing this, side A avoids addressing the real issue. A commonly seen straw man argument is that the theory of evolution is not a valid theory because it does not give a satisfactory explanation for the origin of life. This is a straw man because the theory of evolution does not claim to explain the origin of life – the theory is about what happened to life after it began.
- **Redefining** is to attach a new definition to a term or concept in the middle of a discussion. This is best explained by an example. Person A states: ‘Either it is a living organism with DNA as the genetic material or it is a virus.’ Person B replies: ‘Everything is alive; we do not know everything about viruses.’ This is redefining the central idea in person A’s statement, which is that all living organisms have DNA as their genetic material. Person B is therefore not really addressing the essence of A’s argument.

The use of these types of fallacy is widespread. Note, however, that use of a fallacy does not necessarily make an argument wrong; it just makes it logically invalid. Using fallacies or ignoring established scientific methodologies will result in ‘pseudoscience’.

Pseudoscience

As the name implies, **pseudoscience** is a false form of science (*pseudo* is Greek for ‘false’). Pseudoscience results when biases and fallacies are not avoided, or when the standards of scientific methodology are not adhered to. Homeopathy and acupuncture are examples of pseudoscientific practices that have been shown to have no effect when tested under strict scientific conditions. Intelligent design is a theory that is considered pseudoscientific because it is not testable or falsifiable.

Pseudoscientists sometimes claim that their theories are based on evidence obtained using scientific methodologies. However, closer observations make it clear that their findings cannot be repeated under controlled conditions. Proponents of pseudoscience disqualify this with fallacious arguments, such as ‘the conditions were not exactly the same’, ‘you were cheating’ or ‘I have proof that it does work’. Resisting and rejecting any evidence that challenges its theories distinguishes pseudoscience from true science: scientific theories are constantly tested and adapted if they prove to be wrong. Science is based on evidence, pseudoscience is based on beliefs.

Benveniste’s famous claims about the memory of water as evidence for homeopathy make it clear that even trained scientists can make mistakes that lead to pseudoscience.

The ‘memory of water’ experiment

Jacques Benveniste (1935–2004) was a French immunologist who was at the centre of a major international controversy in 1988, when he published a paper in the scientific journal *Nature*. It described the action of very high dilutions of an antibody on human white blood cells and his findings seemed to support the principle of homeopathy.

Homeopathy is based on the premise that ‘a substance that causes the symptoms of a disease in healthy people will cure similar symptoms in sick people’. In 1796, this form of alternative medicine was first proposed by Samuel Hahnemann (1755–1843). Preparing homeopathic medicines uses repetitive dilutions of a dissolved substance. Dilution factors of 1 in 10^{24} are purported to be effective. At these high dilutions, effectively none of the original (dissolved) substance can be found in solution.

Benveniste used this approach to test if antibodies could still trigger a reaction in human white blood cells. According to the original publication, these very high dilutions still caused an effect. The effect was referred to as the ‘memory of water’.

The claims were highly controversial and, as a condition of publication, *Nature* asked for the results to be replicated by independent laboratories. After the article was published, a follow-up investigation was carried out with the cooperation of Benveniste’s own team. It failed to replicate the original results. Subsequent investigations by other research teams also could not corroborate the original claims. Benveniste refused to retract his controversial article. He claimed that the follow-up investigation had deviated from the original protocols and therefore that these new findings were invalid.



Final note

Science is one of humanity's greatest creations. Its achievements touch every aspect of life and have brought progress to many millions of people. These achievements are the result of collaboration, adherence to strict protocols, persistence, corroboration or falsification of previous findings, and a deep trust of the principles of science.

Many big questions are still unanswered. A large number of the problems we are facing are intercultural or international in scope and epic in size – just think of climate change, the energy crisis, curing cancer, or the rapid evolution of bacteria and viruses. Improving our chances of success requires international coordination and substantial funding, but all these challenges can be conquered.

Option A Relativity

A1 The beginnings of relativity

It is said that Albert Einstein, as a boy, asked himself what would happen if he held a mirror in front of himself and ran forward at the speed of light. With respect to the ground, the mirror would be moving at the speed of light. Rays of light leaving young Einstein's face would also be moving at the speed of light relative to the ground. This meant that the rays would not be moving relative to the mirror, and hence there should be no reflection in it. This seemed odd to Einstein. He expected that looking into the mirror would not reveal anything unusual. Some years later, Einstein would resolve this puzzle with a revolutionary new theory of space and time, the theory of special relativity.

A1.1 Reference frames

In a physics experiment, an observer records the time and position at which **events** take place. To do that, she uses a **reference frame**.

A reference frame is a set of coordinate axes and a set of clocks at every point in space. If this set is not accelerating, the frame is called an **inertial reference frame** (Figure A.1).

So if the 'event' is a lightning strike, an observer will look at the reading of the clock at the point where lightning struck and record that reading as the time of the event. The coordinates of the strike point give the position of the event in space. So, in Figure A.2, lightning strikes at time $t = 3$ s and position $x = 60$ m and $y = 0$. (We are ignoring the z coordinate.)

The same event can also be viewed by another observer in a different frame of reference. Consider, therefore, the following situation involving one observer on the ground and another who is a passenger on a train.

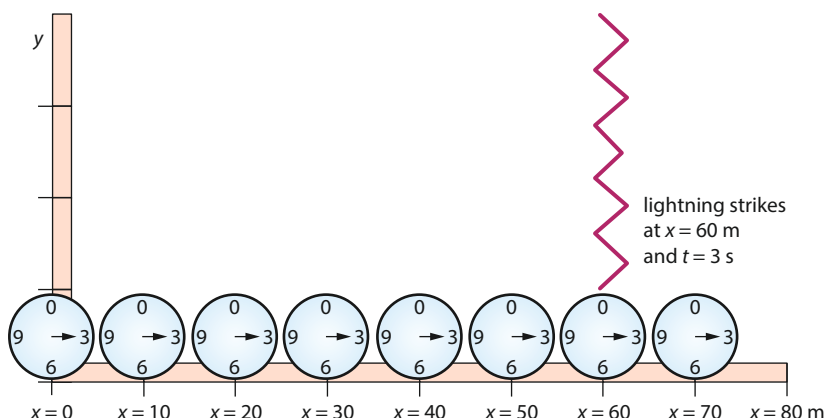


Figure A.2 In this frame of reference the observer decides that lightning struck at time $t = 3$ s and position $x = 60$ m.

Learning objectives

- Use reference frames.
- Understand Galilean relativity with Newton's postulates for space and time.
- Understand the consequences of Maxwell's theory for the speed of light.
- Understand how magnetic effects are a consequence of relativity.

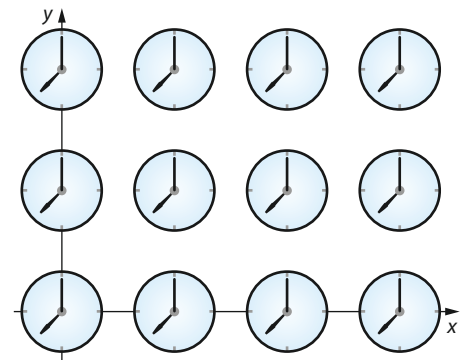


Figure A.1 A two-dimensional reference frame. There are clocks at every point in space (only a few are shown here). All clocks show the same time.

In Figure A.3, a train moves past the observer on the ground (represented by the thin vertical line) at time $t=0$, and is struck by lightning 3 s later. The observer on the train is represented by the thick vertical line. The train is travelling at a velocity of $v=15\text{ m s}^{-1}$ as far as the ground observer is concerned.

Let us see how the two observers view various events along the trip. Assume first that when the clocks carried by the two observers show zero, the origins of the rulers carried by the observers coincide, as shown in Figure A.3. The stationary observer uses the symbol x to denote the distance of an event from the origin of his ruler. The observer on the train uses the symbol x' .

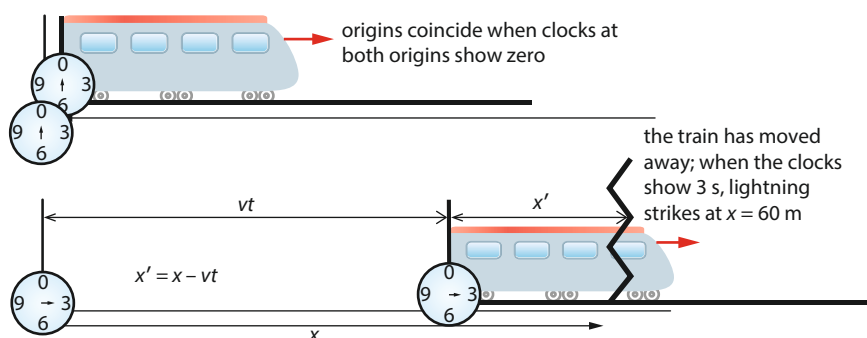


Figure A.3 The origins of the two frames of reference coincide when clocks in both frames show zero. The origins then separate.

Lightning strikes a point on the train. The observer on the train measures the point where the lightning strikes and finds the answer to be x' . The ground observer measures that the lightning struck at a distance x from his origin.

$$x' = x - vt$$

$$t' = t$$

In time t s the train moves forward a distance vt . These equations express the relationship between the coordinates of the same event as viewed by two observers who are in relative motion.

Thus, to the event in Figure A.3 ('lightning strikes') the ground observer assigns the coordinates $x = 60\text{ m}$ and $t = 3\text{ s}$. The observer on the train assigns to this same event the coordinates $t' = 3\text{ s}$ and $x' = 60 - 15 \times 3 = 15\text{ m}$.

We are assuming here what we know from everyday experience (a guide that, as we will see, may not always be reliable): that two observers always agree on what the time coordinates are; in other words, time is common to both observers. Or, as Newton wrote,

Absolute, true and mathematical time, of itself, and from its own nature, flows equably without any relation to anything external.

Of course, the observer on the train may consider herself at rest and the ground below her to be moving away with velocity $-v$. It is impossible for one of the observers to claim that he or she is 'really' at rest and that the other is 'really' moving. There is no experiment that can be performed by, say, the observer on the train that will convince her that she 'really' moves (apart from looking out of the window). If we



consider instead a space station and a spacecraft in outer space as our two frames, even looking out of the window will not help. Whatever results the observer on the train gets out of her experiments, the ground observer also gets out of the same experiments performed in his ground frame of reference. The equations

$$x' = x - vt$$

$$t' = t$$

are called **Galilean transformation** equations, in honour of Galileo (Figure A.4): the relationship between the coordinates of an event when one frame moves past the other with uniform velocity in a straight line. Both observers are equally justified in considering themselves to be at rest, and the descriptions they give are equally valid.

Galilean relativity has an immediate consequence for the **law of addition of velocities**. Consider a ball that rolls with velocity u' as measured by the observer on the train. Again assume that the two frames, train and ground, coincide when $t = t' = 0$ and that the ball first starts rolling when $t' = 0$. Then, after time t' the position of the ball is measured to be at $x' = u't'$ by the observer on the train (Figure A.5).

The ground observer records the position of the ball to be at $x = x' + vt = (u' + v)t$ (recall $t = t'$), so as far as the ground observer is concerned the ball has a velocity (distance/time) given by

$$u = u' + v$$

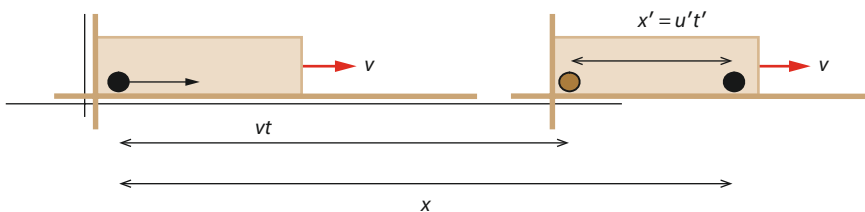


Figure A.5 An object rolling on the floor of the 'moving' frame appears to move faster as far as the ground observer is concerned.

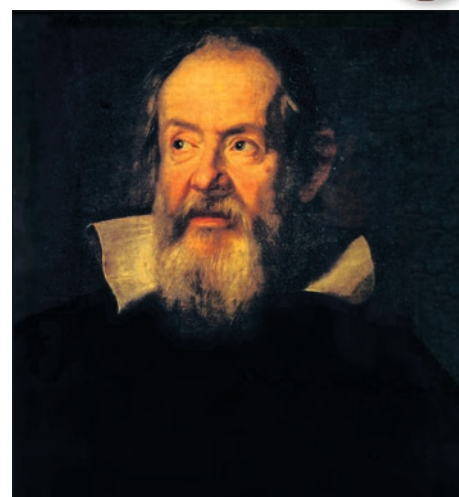


Figure A.4 Galileo Galilei (1564–1642).

Worked example

A.1 A ball rolls on the floor of a train at 2 ms^{-1} (with respect to the floor). The train moves with respect to the ground **a** to the right at 12 ms^{-1} , **b** to the left at 12 ms^{-1} . What is the velocity of the ball relative to the ground?

- a** The velocity is 14 ms^{-1} .
- b** The velocity is -10 ms^{-1} .

This apparently foolproof argument presents problems, however, if we replace the rolling ball in the train by a beam of light moving with velocity $c = 3 \times 10^8 \text{ ms}^{-1}$ as measured by the observer on the train. Using the formula above implies that light would be travelling at a higher speed relative to the ground observer. At the end of the 19th century, considerable efforts were made to detect variations in the speed of light depending on the state of motion of the source of light. The experimental result was that no such variations were detected!

A1.2 Maxwell and the speed of light

In 1864, James Clerk Maxwell corrected an apparent flaw in the laws of electromagnetism by introducing his famous ‘displacement current’ term into the electromagnetic equations. The result is that a changing electric flux produces a magnetic field just as a changing magnetic flux produces an electric field (as Faraday had discovered earlier). An immediate conclusion was that accelerated electric charges produce a pair of self-sustaining electric and magnetic fields at right angles to each other, which eventually decouple from the charge and move away from it at the speed of light. Maxwell discovered **electromagnetic waves** and thus demonstrated the electromagnetic nature of light.

One prediction of the Maxwell theory was that the speed of light is a **universal constant**. Indeed, Maxwell was able to show that the speed of light is given by the expression

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

where the two constants are the electric permittivity and magnetic permeability of free space (vacuum): two constants at the heart of electricity and magnetism.

Maxwell’s theory predicts that the speed of light in vacuum does not depend on the speed of its source.

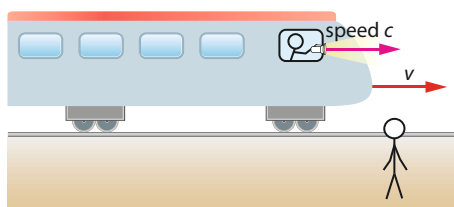


Figure A.6 An observer on the train measures the speed of light to be c . An observer on the ground would then measure a different speed, $c + v$, according to Galileo.

This is in direct conflict with Galilean relativity. According to Galilean relativity, if the speed of light takes on the value c in the ‘train’ frame of reference then it will have the value $c + v$ in the ‘ground’ frame of reference (Figure A.6).

The situation faced by Einstein in 1905 was that the Galilean transformation equations (which were perfectly compatible and consistent with Newtonian mechanics) did not seem to be compatible with Maxwell’s theory. It turned out that, a bit before 1905, the Dutch physicist Hendrik Lorentz (Figure A.7) was also thinking about the same problem.

He too realised that the Maxwell theory was not compatible with the Galilean transformation equations. He set about trying to find the simplest set of transformation equations that would be compatible with Maxwell’s theory. Lorentz’s answer was a strange-looking set of equations, the **Lorentz transformation** equations (Table A.1):

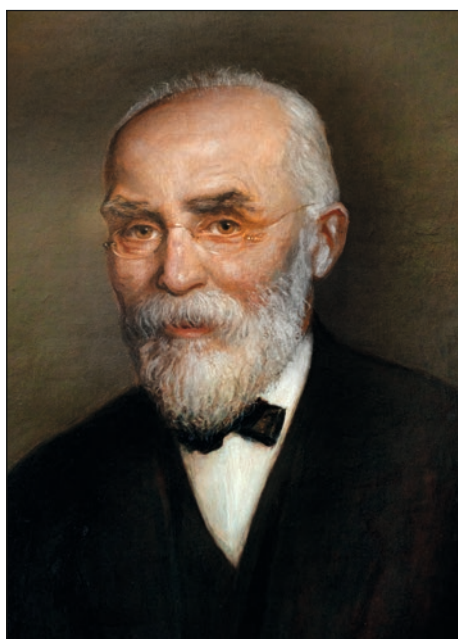


Figure A.7 Hendrik Lorentz (1853–1928).

Galileo	Lorentz
$x' = x - vt$	$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}$
$t' = t$	$t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}}$

Table A.1 Galilean and Lorentz transformations.



As we will soon see, these equations predict two brand new phenomena, which we will discuss in detail later: **length contraction** (a moving object is measured to have a smaller length than when it is at rest) and **time dilation** (moving clocks run slow). But the new problem now was that these new equations were not compatible with Newtonian mechanics! So there were two choices available (Table A.2).

Choice A	Choice B
Accept Newtonian mechanics and the Galilean equations and then modify Maxwell's theory	Accept Maxwell's theory and the Lorentz equations and then modify Newtonian mechanics

Table A.2 The choices Einstein was confronted with.

Einstein's choice was B. He realised early on that all kinds of puzzles arise when one looks at electric and magnetic phenomena from two different inertial reference frames. These could only be resolved if one made choice B.

A1.3 Electromagnetic puzzles

Imagine a wire at rest in a lab, in which a current flows to the left. This means that the electrons in the wire are moving with a drift speed v to the right (Figure A.8). The wire also contains positive charges that are at rest. The average distance between the electrons and that between the positive charges are the same. The net charge of the wire is zero, so an observer at rest in the lab measures zero electric field around the wire. Now imagine a positive charge q that moves with the same velocity as the electrons in the rod. The observer in the lab has no doubt that there will be a **magnetic** force on this charge. The magnetic field at the position of the charge is out of the page and so there will be a magnetic force repelling the charge q from the wire (use the right-hand rule for magnetic force).

Now consider things from the point of view of another observer who moves along with the charge q . For this observer, the charge q and the electrons in the rod are at rest, but the positive charges in the rod move to the left with speed v . These positive charges do produce a magnetic field, but the magnetic force on q is zero since $F = qvB$ and the speed of the charge is zero.

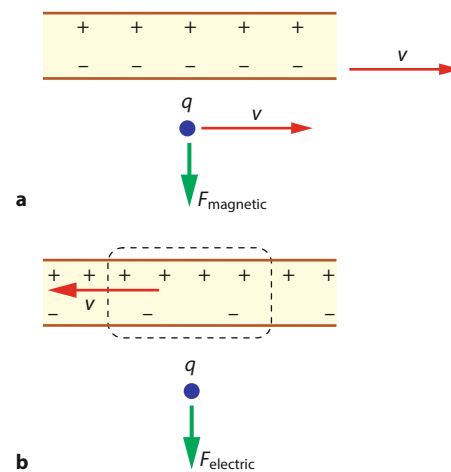


Figure A.8 **a** An observer at rest with respect to the wire measures a magnetic force on the moving charge. **b** An observer moving along with the charge will measure an electric force.



Relativity and reality

Relativity says that the details of an experiment may appear different to different observers in relative motion. This does not mean that different observers experience a different 'reality'. The physically significant aspects of the experiment are the same for both observers.

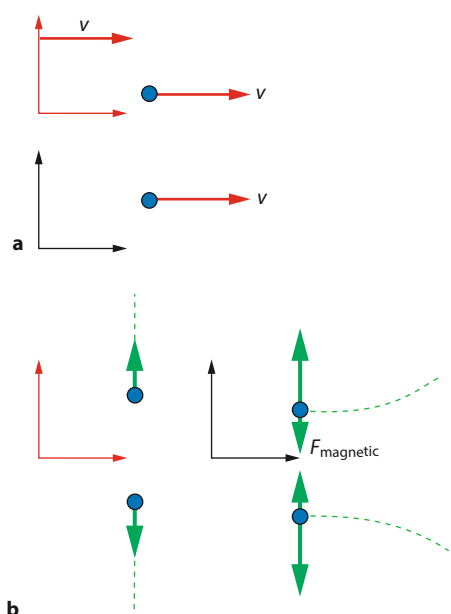


Figure A.9 **a** Two protons move past an observer with the same velocity v . **b** The electric and magnetic forces on the charges look different in each reference frame.

Now, if the charge q feels a force according to one observer, other observers must reach similar conclusions. So the puzzle is: where does the force on the charge q come from? Notice that in Figure A.8b we have made the distance between positive charges smaller and the distance between electrons larger. This has to be the case if Lorentz is right about length contraction. The separation between the positive charges is a distance that moves past the observer and so has to get smaller. The separation between the electrons used to be a ‘moving’ distance so, now that it has stopped, it has to be bigger. The effect of this is that now, as far as the charge q is concerned, there is more positive charge than negative charge in the wire near it. There will then be an **electric** force of repulsion. We have learned that, as a result of length contraction, the force that an observer calls magnetic in one reference frame may be an electric force in another frame.

There are many such puzzles which can only be resolved if the phenomena of length contraction and time dilation are taken into account. So consider two positive charges that move parallel to each other with speed v relative to the black frame (the lab); see Figure A.9a.

For an observer moving along with the charges (red frame) the charges are at rest and so there cannot be a magnetic force between them. There is, however, an electric force of repulsion. For this observer, the charges move away from each other along straight lines. For an observer at rest in the lab there are repulsive electric forces but there are also magnetic forces; this is because each moving charge creates a magnetic field and the other charge moves in that magnetic field. The magnetic field created by the bottom charge at the position of the top charge is directed out of the page. The magnetic force on the top charge is therefore directed opposite to the electric force; the magnetic force is attractive. The net force is still repulsive and the two charges move away from each other along curved paths. Examining the details of this situation shows that the two observers will reach consistent results only if time runs differently in the two different frames. This is more evidence that, to avoid these electromagnetic puzzles, ideas similar to Lorentz’s must be true; it is not a surprise that Einstein’s 1905 paper is entitled ‘On the electrodynamics of moving bodies’.

Worked example

A.2 A positive electric charge q enters a region of magnetic field B with speed v (Figure A.10). Discuss the forces, if any, that the charge experiences according to **a** a frame of reference at rest with respect to the magnetic field (black) and **b** a frame moving with the same velocity as the charge (red).

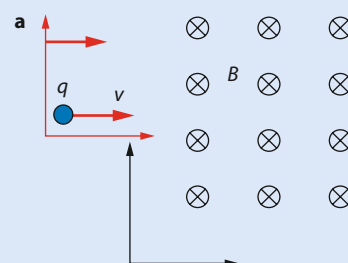


Figure A.10

- a** The charge will experience a magnetic force given by $F = qvB$.
- b** The charge is at rest. Hence the magnetic force is zero. But there has to be a force, and that can only be an electric force.



Nature of science

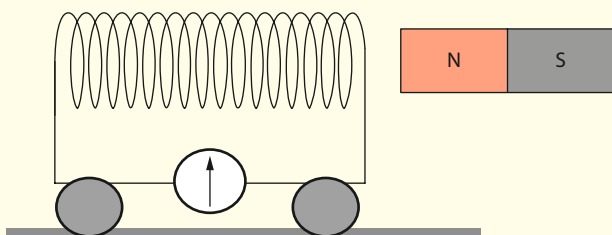
Paradigm shift

A fundamental fact of relativity theory is that the speed of light is constant for all inertial reference frames. This simple postulate has far-reaching consequences for our understanding of space and time. The idea that time is an absolute, assumed to be correct for over 2000 years, was shown to be false. Accepting the new ideas was not easy, but they offered the best explanation of observations and revolutionised physics.

? Test yourself

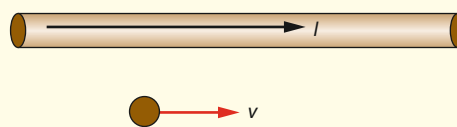
- 1 It was a very hotly debated subject centuries ago as to whether the Earth goes around the Sun or the other way around. Does relativity make this whole argument irrelevant since ‘all frames of reference are equivalent’?
- 2 Discuss the approximations necessary in order to claim that the rotating Earth is an inertial frame of reference.
- 3 Imagine that you are travelling in a train at constant speed in a straight line and that you cannot look at or communicate with the outside. Think of the first experiment that comes to your mind that you could do to try to find out that you are indeed moving. Then analyse it carefully to see that it will not work.

- HL** 4 Here is another experiment that could be performed in the hope of determining whether you are moving or not. The coil shown below is placed near a strong magnet, and a galvanometer attached to the coil registers a current. Discuss whether we can deduce that the coil moves with respect to the ground.



- 5 Outline an experiment you might perform in a train that is accelerating along a straight line that would convince you that it is accelerating. Discuss whether you could also determine the direction of the acceleration.

- 6 Outline how you would know that you find yourself in a rotating frame of reference.
- 7 An electric current flows in a wire. A proton moving parallel to the wire will experience a magnetic force due to the magnetic field created by the current. From the point of view of an observer travelling along with the proton, the proton is at rest and so should experience no force. Analyse the situation.



- 8 An inertial frame of reference S' moving to the right with speed 15 ms^{-1} moves past another inertial frame S . At time zero the origins of the two frames coincide. **a** A phone rings at location $x = 20 \text{ m}$ and $t = 5.0 \text{ s}$. Determine the location of this event in the frame S' . **b** A ball is measured to have velocity 5.0 ms^{-1} in frame S' . Calculate the velocity of the ball in frame S .
- 9 An inertial frame of reference S' moving to the left with speed 25 ms^{-1} moves past another inertial frame S . At time zero the origins of the two frames coincide. **a** A phone rings at location $x' = 24 \text{ m}$ and $t = 5.0 \text{ s}$. Determine the location of this event in the frame S . **b** A ball is measured to have velocity -15 ms^{-1} in frame S . Calculate the velocity of the ball in frame S' .

Learning objectives

- Understand the two postulates of special relativity.
- Describe clock synchronisation.
- Understand and use the Lorentz transformations.
- Apply the velocity addition formula.
- Work with invariant quantities.
- Understand and apply time dilation and length contraction.
- Use muon decay as evidence for relativity.

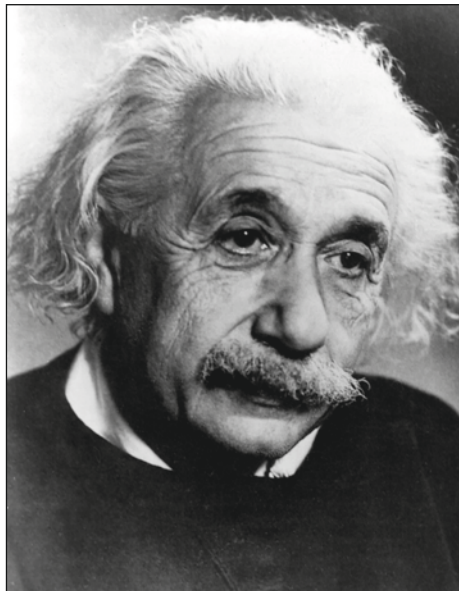


Figure A.11 Albert Einstein (1879–1955).

A2 The Lorentz transformations

This section introduces the postulates of the theory of relativity and the transformation equations that relate the coordinates of an event as seen from two different inertial frames. We discuss the two major consequences of these equations: time dilation and length contraction. We also discuss the **muon decay** experiment, which provides evidence in support of the theory of relativity.

A2.1 The postulates of special relativity

As discussed in Section A1, the Lorentz transformation equations are the simplest set of transformations with which the Maxwell theory is consistent when looked at from different reference frames. Einstein (Figure A.11) re-derived the same set of equations based on two much simpler and more general assumptions. These two assumptions are known as the **postulates of relativity**. They are:

- The laws of physics are the same in all inertial frames.
- The speed of light in vacuum is the same for all inertial observers.

So we see that it is not just electromagnetism that must be consistent with these transformations, but all the laws of physics.

These two postulates of relativity, although they sound simple, have far-reaching consequences. The fact that the speed of light in a vacuum is the same for all observers means that absolute time does not exist. Consider a beam of light. Two different observers in relative motion to each other will measure different distances covered by this beam. But if they are to agree that the speed of the beam is the same for both observers, it follows that they must also measure different times of travel. Thus, observers in motion relative to each other measure time differently. The constancy of the speed of light means that space and time are now inevitably linked and are not independent of each other, as they were in Newtonian mechanics.



The strength of intuition

Einstein was so convinced of the constancy of the speed of light that he elevated it to one of the principles of relativity. But it was not until 1964 that conclusive experimental verification of this took place. In an experiment at CERN, neutral pions moving at $0.99975c$ decayed into a pair of photons moving in different directions. The speed of the photons in both directions was measured to be c with extraordinary accuracy. The speed of light does not depend on the speed of its source.



We have seen that Galileo's transformation laws

$$x' = x - vt, \quad t' = t, \quad u' = u - v$$

need to be changed. Einstein (and Lorentz) modified them to

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Using the **gamma factor**, $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$, we may rewrite these as

$$x' = \gamma(x - vt), \quad t' = \gamma\left(t - \frac{v}{c^2}x\right)$$

These formulas are useful when we know x and t and we want to find x' and t' . If, on the other hand, we know x' and t' and want to find x and t , then

$$x = \gamma(x' + vt'), \quad t = \gamma\left(t' + \frac{v}{c^2}x'\right)$$

These equations may be used to relate the coordinates of a **single** event in one inertial frame to those in another. We will denote by S some inertial frame. A frame that moves with speed v to the right relative to S we will call S' . These equations assume that, when clocks in both frames show zero (i.e. $t = t' = 0$), the origins of the two frames coincide (i.e. $x = x' = 0$).

Note that $\gamma > 1$. A graph of the gamma factor γ versus velocity is shown in Figure A.12. We see that γ is approximately 1 for velocities up to about half the speed of light, but approaches infinity as the speed approaches the speed of light.

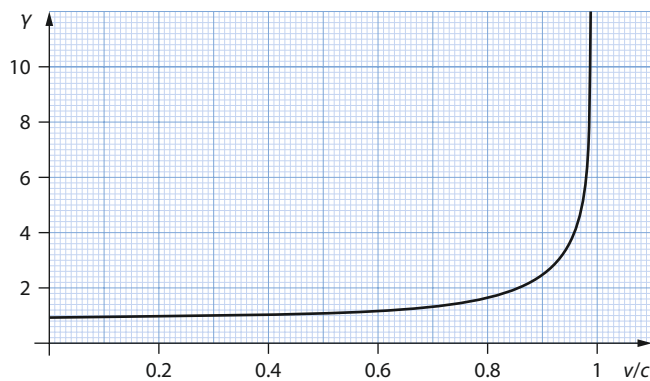


Figure A.12 The gamma factor γ as a function of velocity. The value of γ stays essentially close to 1 for values of the velocity up to about half the speed of light, but approaches infinity as the velocity approaches the speed of light.

A2.2 Clock synchronisation

The equations above also assume that the clocks in each frame are synchronised. This means that they show the same time at any given instant of time, i.e. we have **clock synchronisation**. How can this be achieved? The idea is to use the fact that the speed of light is a universal constant. Consider a clock that is a distance x from the origin. A light signal leaving the origin at time zero will take a time $\frac{x}{c}$ to arrive at the clock.

So we set that clock to show a time equal to $\frac{x}{c}$ and wait for a light signal from the origin to arrive. When the light signal does arrive, the clock is started (the clock is a stopwatch). Doing this for all the clocks in the reference frame ensures that they are all synchronised.

Worked examples

A.3 Lightning strikes a point on the ground (frame S) at position $x = 3500$ m and time $t = 5.0$ s. Determine where and when the lightning struck according to a rocket that flies to the right over the ground at a speed of $0.80c$. (Assume that when clocks in both frames show zero, the origins of the two frames coincide.)

This is a straightforward application of the Lorentz formulas (the gamma factor is $\gamma = \frac{1}{\sqrt{1-0.80^2}} = \frac{5}{3}$):

$$\begin{aligned}x' &= \gamma(x - vt) \\ &= \frac{5}{3}(3500 - 0.80 \times 3 \times 10^8 \times 5.0) \\ &= -2.0 \times 10^9 \text{ m}\end{aligned}$$

and

$$\begin{aligned}t' &= \gamma\left(t - \frac{vx}{c^2}\right) \\ &= \frac{5}{3}\left(5.0 - \frac{0.80 \times 3 \times 10^8 \times 3500}{(3 \times 10^8)^2}\right) \\ &= 8.3 \text{ s}\end{aligned}$$

The observers disagree on the coordinates of the event 'lightning strikes'; both are correct.

A.4 A clock in a rocket moving at $0.80c$ goes past a lab. When the origins of both frames coincide, all clocks are set to show zero. Calculate the reading of the rocket clock as it goes past the point with $x = 240$ m.

The event 'clock goes past the given point' has coordinates x and t in the lab frame and x' and t' in the rocket frame. We need to find t' . We know that $x = 240$ m. We calculate that $t = \frac{240}{0.80c} = 1.0 \times 10^{-6}$ s. Hence, from the Lorentz formula for time.

$$\begin{aligned}t' &= \gamma\left(-\frac{vx}{c^2}\right) \\ &= \frac{5}{3}\left(1.0 \times 10^{-6} - \frac{0.80 \times 3 \times 10^8 \times 240}{(3 \times 10^8)^2}\right) \\ &= 0.60 \mu\text{s}\end{aligned}$$

Exam tip

It is important that you are able to use the Lorentz equations to go from frame S to frame S' as well as from S' to S.



Sometimes we will be interested in the difference in coordinates of a pair of events. So, for events 1 and 2, we define

$$\Delta x' = x_2' - x_1' \text{ and } \Delta t' = t_2' - t_1' \text{ in } S'$$

and

$$\Delta x = x_2 - x_1 \text{ and } \Delta t = t_2 - t_1 \text{ in } S$$

These differences are related by

$$\Delta x' = \gamma(\Delta x - v\Delta t); \Delta t' = \gamma\left(\Delta t - \frac{v}{c^2}\Delta x\right)$$

or, in reverse,

$$\Delta x = \gamma(\Delta x' + v\Delta t'); \Delta t = \gamma\left(\Delta t' + \frac{v}{c^2}\Delta x'\right)$$

Worked examples

A.5 Consider a rocket that moves relative to a lab with speed $v = 0.80c$ to the right. We call the inertial frame of the lab S and that of the rocket S' . The length of the rocket as measured by an observer on the rocket is 630 m. A photon is emitted from the back of the rocket towards the front end. Calculate the time taken for the photon to reach the front end of the rocket according to observers in the rocket and in the lab.

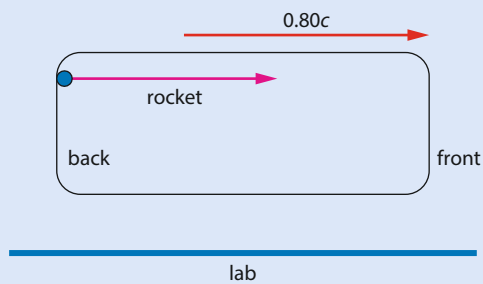


Figure A.13

Event 1 is the emission of the photon. Event 2 is the arrival of the photon at the front of the rocket. Clearly,

$$\Delta x' = x'_{\text{arrival}} - x'_{\text{emission}} = 630 \text{ m}$$

This distance is covered at the speed of light and so the time between emission and arrival is

$$\Delta t' = \frac{630}{c} = 2.1 \times 10^{-6} \text{ s}$$

In frame S we use

$$\Delta x = \frac{5}{3}(630 + 0.80c \times 2.1 \times 10^{-6}) = 1890 \text{ m}$$

$$\Delta t = \frac{5}{3}\left(2.1 \times 10^{-6} + \frac{0.80c \times 630}{c^2}\right) = 6.30 \times 10^{-6} \text{ s}$$

A.6 a A rocket approaches a space station with speed $0.980c$ (relative to the space station). Observers in the rocket record the firing of a missile from the space station and 1.0 s later (according to rocket clocks) record an explosion at another space station behind the rocket at a distance of $4.0 \times 10^8\text{ m}$ (also measured according to the rocket). Calculate the distance and the time interval between the events ‘firing of the missile’ and ‘explosion’ according to observers in the space station. Comment on your answers.

b In frame S , event A causes event B and therefore occurs before event B . Show that in any frame, event A always occurs before event B .

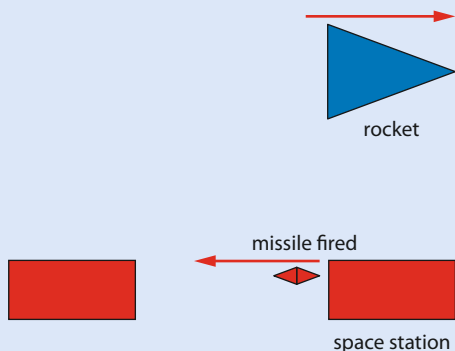


Figure A.14

a The gamma factor for a speed of $0.980c$ is $\gamma = \frac{1}{\sqrt{1-0.980^2}} \approx 5.0$. We will take the space station as frame S and the rocket as frame S' . Then we know that $\Delta x' = x'_{\text{expl}} - x'_{\text{fire}} = -4.0 \times 10^8\text{ m}$; $\Delta t' = t'_{\text{expl}} - t'_{\text{fire}} = 1.0\text{ s}$. The explosion happens after the firing of the missile and so $\Delta t' > 0$. We apply the (reverse) Lorentz transformations to find

$$\Delta x = 5.0 \times (-4.0 \times 10^8 + 0.98c \times 1.0) = -5.3 \times 10^8\text{ m}$$

$$\Delta t = 5.0 \times \left(1.0 + \frac{0.98c \times (-4.0 \times 10^8)}{c^2} \right) = -1.5\text{ s}$$

The negative answer for the time interval means that the explosion happens **before** the firing of the missile. Therefore the missile could not have been responsible for the explosion! We could have predicted that the missile had nothing to do with the explosion because, according to observers in the rocket, the missile would have to cover a distance of $4.0 \times 10^8\text{ m}$ in 1.0 s , that is, with a speed of $4.0 \times 10^8\text{ m s}^{-1}$, which exceeds the speed of light and so is impossible.

b We are told that $\Delta t = t_B - t_A > 0$. Let Δx be the distance between the two events. The fastest way information from A can reach B is at the speed of light, and so $\frac{\Delta x}{\Delta t} < c$. In any other frame,

$$\begin{aligned} \Delta t' &= t'_B - t'_A \\ &= \gamma \left(\Delta t - \frac{v \Delta x}{c^2} \right) \\ &= \gamma \Delta t \left(1 - \frac{v \Delta x}{c^2 \Delta t} \right) \\ &> \gamma \Delta t \left(1 - \frac{vc}{c^2} \right) = \gamma \Delta t \left(1 - \frac{v}{c} \right) \\ &> 0 \end{aligned}$$

So event A has to occur before event B in any frame.



A2.3 Time dilation

Suppose that an observer in frame S' measures the time between successive ticks of a clock. The clock is at rest in S' and so the measurement of the ticks occurs at the same place, $\Delta x' = 0$. The result of his measurement is $\Delta t'$. The observer in S will measure a time interval equal to

$$\begin{aligned}\Delta t &= \gamma \left(\Delta t' - \frac{v \Delta x'}{c^2} \right) \\ &= \gamma (\Delta t' + 0) \\ \Delta t &= \gamma \Delta t'\end{aligned}$$

This shows that the interval between the ticks of a clock that is moving relative to the observer in S is greater than the interval measured in S' where the clock is at rest. This is known as **time dilation**. In this case, the time measured in S' is special because it represents a time interval between two events that happened at the same point in space. The clock is at rest in S' so its ticks happen at the same point in S' . Such a time interval is called **proper time interval**. (This is just a name; it is not implied that this is a more 'correct' measurement of time.)

A proper time interval is the time in between two events that take place at the same point in space.

The time dilation formula is best remembered as

time interval = γ × proper time interval

$$= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \times \text{proper time interval}$$

Exam tip

Do not make the mistake of thinking that proper time intervals are measured only in the frame S' . A proper time interval is the time between two events that take place at the same point.



Different observers disagree in their measurements, but both are right

Note that there is no question as to which observer is right and which is wrong when it comes to measuring time intervals. Both are right. Two inertial observers moving relative to each other at constant velocity both reach valid conclusions, according to the principle of relativity.

Worked examples

A.7 The time interval between the ticks of a clock carried on a fast rocket is half of what observers on the Earth record. Calculate the speed of the rocket relative to the Earth.

From the time dilation formula it follows that

$$\begin{aligned}2 &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \\ \Rightarrow 1 - \frac{v^2}{c^2} &= \frac{1}{4} \\ \Rightarrow \frac{v^2}{c^2} &= \frac{3}{4} \\ \Rightarrow v &= 0.866c\end{aligned}$$

A.8 A rocket moves past an observer in a laboratory with speed $v = 0.85c$. An observer in the laboratory measures that a radioactive sample of mass 50 mg (which is at rest in the laboratory) has a half-life of 2.0 min. Calculate the half-life as measured by observers in the rocket.

We have two events here. The first is that the laboratory observer sees a container with 50 mg of the radioactive sample. The second event is that the laboratory observer sees a container with 25 mg of the radioactive sample. These events are separated by 2.0 min as far as the laboratory observer is concerned. These two events take place at the same point in space as far as the laboratory observer is concerned, so the laboratory observer has measured the **proper time interval** between these two events. Hence the rocket observers will measure a longer half-life, equal to $\gamma \times$ (proper time interval)

$$\begin{aligned} &= \frac{1}{\sqrt{1-0.85^2}} \times 2.0 \text{ min} \\ &= 3.80 \text{ min} \end{aligned}$$

The point of this example is that you must not make the mistake of thinking that proper time intervals are measured by 'the moving' observer. There is no such thing as 'the moving' observer: the rocket observer is free to consider herself at rest and the laboratory observer to be moving with velocity $v = -0.85c$.

Exam tip

Note that it is only lengths in the direction of motion that contract.

A2.4 Length contraction

Now consider a rod that is at rest in frame S' . An observer in S' measures the position of the ends of the rod, subtracts and finds a length $\Delta x' = L_0$. (Notice that, since the rod is at rest, the measurements of the position of the ends can be done at different times: the rod is not going anywhere). But for an observer in S the rod is moving, so to measure its length he must record the position of the ends of the rod at the same time, that is, with $\Delta t = 0$. Since $\Delta x' = \gamma(\Delta x - v\Delta t)$, we obtain the result that

$$L_0 = \gamma(L - 0)$$

$$L = \frac{L_0}{\gamma}$$

This shows that the rod, which is moving relative to the observer in S , has a shorter length in S than the length measured in S' , where the rod is at rest. This is known as **length contraction**.

The length $\Delta x' = L_0$ is special because it is measured in a frame of reference where the rod is at rest. Such a length is called a **proper length**.

The length of an object measured by an inertial observer with respect to whom the object is at rest is called a proper length. Observers with respect to whom the object moves at speed v measure a shorter length:

$$\text{length} = \frac{\text{proper length}}{\gamma}$$



We must accept experimentally verified observations, no matter how 'strange' they may appear

Time dilation as described is a 'real' effect. In the Hafele–Keating experiment, accurate atomic clocks taken for a ride aboard planes moving at ordinary speeds and then compared with similar clocks left behind show readings that are smaller by amounts consistent with the formulas of relativity. Time dilation is also a daily effect in the operation of particle accelerators. In such machines, particles are accelerated to speeds that are very close to the speed of light and thus relativistic effects must be taken into account when designing these machines. The time dilation formula has also been verified in muon– decay experiments.

Worked example

A.9 In the year 2014, a group of astronauts embark on a journey towards the star Betelgeuse in a spacecraft moving at $v = 0.75c$ with respect to the Earth. Three years after departure from the Earth (as measured by the astronauts' clocks) one of the astronauts announces that she has given birth to a baby girl. The other astronauts immediately send a radio signal to the Earth announcing the birth. Calculate when the good news is received on the Earth (according to Earth clocks).

When the astronaut gives birth, three years have gone by according to the spacecraft's clocks. This is a proper time interval since the events 'departure from Earth' and 'astronaut gives birth' happen at the same place as far as the astronauts are concerned (inside the spacecraft). Thus, the time between these two events according to the Earth clocks is

$$\begin{aligned}\text{time interval} &= \gamma \times \text{proper time interval} \\ &= \frac{1}{\sqrt{1-0.75^2}} \times 3.0 \text{ yr} \\ &= 4.54 \text{ yr}\end{aligned}$$

This is therefore also the time during which the spacecraft has been travelling, as far as the Earth is concerned. The distance covered is (as far as the Earth is concerned)

$$\begin{aligned}\text{distance} &= vt \\ &= 0.75c \times 4.54 \text{ yr} \\ &= 3.40c \text{ yr} \\ &= 3.40 \text{ ly}\end{aligned}$$

Exam tip

$$1 \text{ ly} = c \times 1 \text{ year}$$

This is the distance that the radio signal must then cover in bringing the message. This is done at the speed of light and so the time taken is

$$\begin{aligned}\frac{340 \text{ ly}}{c} &= \frac{3.40c \times \text{yr}}{c} \\ &= 3.40 \text{ yr}\end{aligned}$$

Hence, when the signal arrives, the year on the Earth is $2014 + 4.54 + 3.40 = 2022$ (approximately).

A2.5 Another look at time dilation

We have seen using Lorentz transformations that relativity predicts the phenomenon of time dilation. A simpler view of this effect is the following. A direct consequence of the principle of relativity is that observers who are in motion relative to each other do not agree on the interval of time separating two events. To see this, consider the following situation: a train moves with velocity v with respect to the ground as shown in Figure A.15.

From point A on the train floor, a light signal is sent towards point B directly above on the ceiling. The time $\Delta t'$ it takes for light to travel from A to B and back to A is recorded. Note that as far as the observers inside the train are concerned, the light beam travels along the straight-line segments AB and BA. From the point of view of an observer on the ground, however, things look somewhat different. In the time it takes for light to return to A, point B (which moves along with the train) will have moved forward. This means, therefore, that to this observer the path of the light beam looks like that shown in Figure A.16.

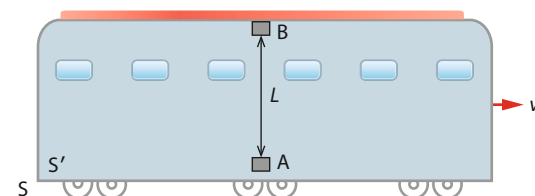


Figure A.15 A signal is emitted at A, is reflected off B and returns to A again. The path shown is what the observer on the train sees as the path of the signal. The emission and reception of the signals happen at the same point in space.

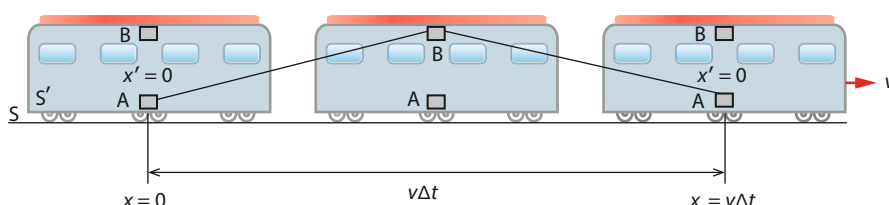


Figure A.16 The ground observer sees things differently. In the time it takes the signal to return, the train has moved forward. Thus, the emission and reception of the signal do not happen at the same point in space. (The diagram is exaggerated for clarity.)

Suppose that the stationary observer measures the time it takes light to travel from A to B and back to A to be Δt . What we will show now is that, if both observers are to agree that the speed of light is the same, then the two time intervals cannot be the same.

It is clear that

$$\Delta t' = \frac{2L}{c}$$

$$\Delta t = \frac{2\sqrt{L^2 + (v\Delta t/2)^2}}{c}$$



since the train moves forward a distance $v\Delta t$ in the time interval Δt that the stationary observer measures for the light beam to reach A. Thus, solving the first equation for L and substituting in the second (after squaring both equations)

$$(\Delta t)^2 = \frac{4\left(\frac{c^2(\Delta t')^2}{4} + \frac{v^2(\Delta t)^2}{4}\right)}{c^2}$$

$$c^2(\Delta t)^2 = c^2(\Delta t')^2 + v^2(\Delta t)^2$$

$$(c^2 - v^2)(\Delta t)^2 = c^2(\Delta t')^2$$

$$(\Delta t)^2 = \frac{c^2(\Delta t')^2}{c^2 - v^2}$$

So, finally,

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{v^2}{c^2}}}$$

This is the **time dilation** formula. It is customary to call the expression

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

the gamma factor, in which case $\Delta t = \gamma\Delta t'$. Recall that $\gamma > 1$.

The time interval for the travel of the light beam is longer on the ground observer's clock. This is known as time dilation. If the train passengers measure a time interval of $\Delta t' = 6.0$ s and the train moves at a speed $v = 0.80c$, then the time interval measured by the ground observer is

$$\begin{aligned} \Delta t &= \frac{6.0}{\sqrt{1 - 0.80^2}} \text{ s} \\ &= \frac{6.0}{\sqrt{1 - 0.64}} \text{ s} \\ &= \frac{6.0}{\sqrt{0.36}} \text{ s} \\ &= \frac{6.0}{0.6} \text{ s} \\ &= 10 \text{ s} \end{aligned}$$

which is longer. It will be seen immediately that this large difference came about only because we chose the speed of the train to be extremely close to the speed of light. Clearly, if the train speed is small compared with the speed of light, then $\gamma \approx 1$, and the two time intervals agree, as we might expect them to from everyday experience. The reason that our everyday experience leads us astray is because the speed of light is enormous compared with everyday speeds. Thus, the relativistic time dilation effect we have just discovered becomes relevant only when speeds close to the speed of light are encountered.

A2.6 Addition of velocities

Consider a frame S' (for example, a train) that moves at constant speed v in a straight line relative to another frame S (for example, the ground). An object slides on the train floor in the same direction as the train (S') and its velocity is measured, by observers in S' , to be u' . What is the speed u of this object as measured by observers in S ? (See Figure A.17.)

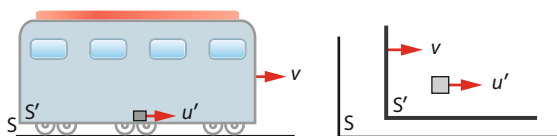


Figure A.17 The speed of the moving object is u' in the frame S' . What is its speed as measured in frame S ?

In pre-relativity physics (i.e. Galilean–Newtonian physics), the answer would be simply $u' + v$. This cannot, however, be the correct relativistic answer; if we replaced the sliding object by a beam of light ($u' = c$), we would end up with an observer (S) who measured a speed of light different from that measured by S' . The correct answer for the speed u of the particle relative to S is

$$u = \frac{u' + v}{1 + \frac{u'v}{c^2}}$$

or, solving for u' ,

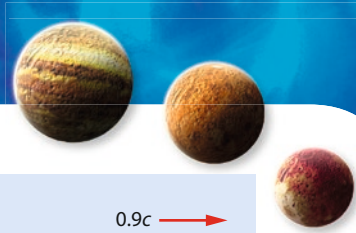
$$u' = \frac{u - v}{1 - \frac{uv}{c^2}}$$

It can easily be checked that, irrespective of how close u' or v is to the speed of light, u is always less than c . For the case in which $u' = c$, then $u = c$ as well, as demanded by the principle of the constancy of the speed of light (check this). On the other hand, if the velocities involved are small compared with the speed of light, then we may neglect the term $\frac{u'v}{c^2}$ in the denominator, in which case Einstein's formula reduces to the familiar Galilean relativity formula $u = u' + v$.

Worked examples

A.10 An electron has a speed of $2.00 \times 10^8 \text{ ms}^{-1}$ relative to a rocket, which itself moves at a speed of $1.00 \times 10^8 \text{ ms}^{-1}$ with respect to the ground. What is the speed of the electron with respect to the ground?

Applying the formula above with $u' = 2.00 \times 10^8 \text{ ms}^{-1}$ and $v = 1.00 \times 10^8 \text{ ms}^{-1}$, we find $u = 2.45 \times 10^8 \text{ ms}^{-1}$.



A.11 Two rockets move away from each other with speeds of $0.8c$ and $0.9c$ with respect to the ground, as shown in Figure A.18. What is the speed of each rocket as measured from the other? What is the relative speed of the two rockets as measured from the ground?



Figure A.18

Let us first find the speed of A with respect to B. In the frame of reference in which B is at rest, the ground moves to the left with speed $0.90c$ (i.e. a velocity of $-0.90c$). The velocity of A with respect to the ground is $-0.80c$. This is illustrated in Figure A.19.

Applying the formula with the values given in the figure, we find

$$\begin{aligned}
 u &= \frac{u' + v}{1 + \frac{u'v}{c^2}} \\
 &= \frac{-0.90c - 0.80c}{1 + \frac{(-0.90c)(-0.80c)}{c^2}} \\
 &= \frac{-1.70c}{1 + 0.72} \\
 &= -\frac{1.70c}{1.72} \\
 &= -0.988c \approx -0.99c
 \end{aligned}$$

The minus sign means, of course, that rocket A moves to the left relative to B. Let us now find the speed of rocket B as measured in A's **rest frame**. The appropriate diagram is shown in Figure A.20.

We thus find

$$\begin{aligned}
 u &= \frac{u' + v}{1 + \frac{u'v}{c^2}} \\
 &= \frac{0.90c + 0.80c}{1 + \frac{(0.90c)(0.80c)}{c^2}} \\
 &= \frac{1.70c}{1 + 0.72} \\
 &= \frac{1.70c}{1.72} \\
 &= 0.988c \approx 0.99c \text{ as expected.}
 \end{aligned}$$

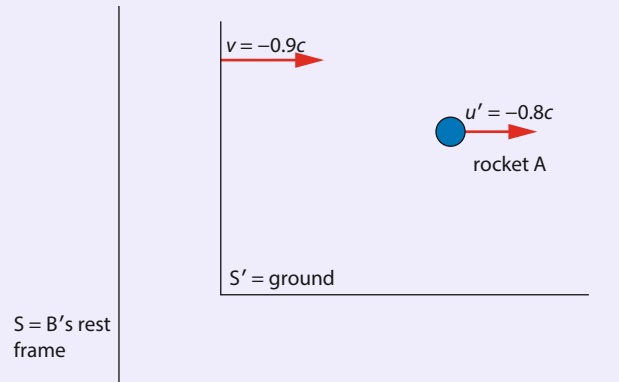


Figure A.19

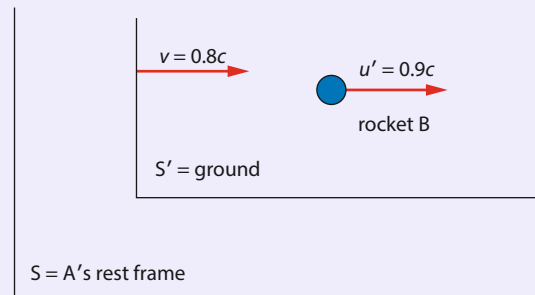


Figure A.20

A.12 A rocket has a proper length of 250 m and travels at a speed $v = 0.950c$ relative to the Earth. A missile is fired from the back of the rocket at a speed $u' = 0.900c$ relative to the rocket. **a** Calculate the time when the missile passes the front of the rocket according to **i** observers in the rocket and **ii** according to observers on the Earth. **b** Calculate the speed of the missile relative to the Earth and show that your answer is consistent with the answer in **a**.

a i Calling, as usual, the Earth frame S and the rocket frame S' , we know that $\Delta x' = 250$ m and

$$\Delta t' = \frac{250}{0.900 \times 3.0 \times 10^8} = 9.259 \times 10^{-7} \approx 9.26 \times 10^{-7} \text{ s, where we are considering the events 'missile is fired' and 'missile passes the front of the rocket'.$$

ii We need to find Δt . From the Lorentz formulas we know that $\Delta t = \gamma \left(\Delta t' + \frac{v \Delta x'}{c^2} \right)$. The gamma factor is

$$\gamma = \frac{1}{\sqrt{1 - 0.950^2}} = 3.20. \text{ Hence}$$

$$\begin{aligned} \Delta t &= 3.20 \times \left(9.259 \times 10^{-7} + \frac{0.950 \times 250}{3.0 \times 10^8} \right) \\ &= 5.496 \times 10^{-6} \approx 5.50 \times 10^{-6} \text{ s} \end{aligned}$$

b From the relativistic addition law for velocities,

$$\begin{aligned} u &= \frac{u' + v}{1 + \frac{u'v}{c^2}} \\ &= \frac{0.900c + 0.950c}{1 + \frac{(0.900c)(0.950c)}{c^2}} \\ &= 0.997c \end{aligned}$$

The distance travelled by the missile according to Earth observers is

$$\begin{aligned} \Delta x &= \gamma(\Delta x' + v \Delta t') \\ &= 3.20 \times (250 + 0.950 \times 3.0 \times 10^8 \times 9.26 \times 10^{-7}) \\ &= 1645 \text{ m} \end{aligned}$$

This distance was covered at a speed of $0.997c$ and so the time taken, according to Earth observers, is

$$\frac{1645}{0.997 \times 3.0 \times 10^8} = 5.50 \times 10^{-6} \text{ s, as expected from a.}$$



A2.7 Simultaneity

Another great change introduced into physics as a result of relativity is the concept of **simultaneity**. Imagine three rockets A, B and C travelling with the same constant velocity (with respect to some inertial observer) along the same straight line. Imagine that rocket B is halfway between rockets A and C, as shown in Figure A.21.

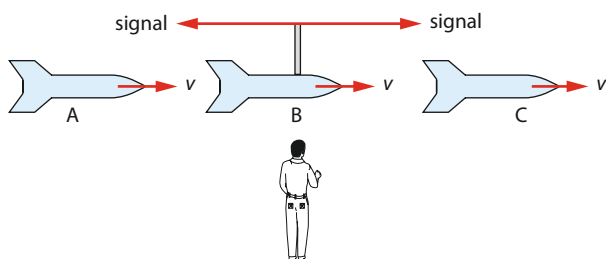


Figure A.21 B emits signals to A and C. These are received at the same time as far as B is concerned. But the reception of the signals is not simultaneous as far as the ground observer is concerned.

Rocket B emits light signals that are directed towards rockets A and C. Which rocket will receive the signal first? The principle of relativity allows us to determine that, as far as an observer in rocket B is concerned, the signals are received by A and C simultaneously (i.e. at the same time). This is obvious, since we may imagine a big box enclosing all three rockets that moves with the same velocity as the rockets themselves. Then, for any observer in the box, or in the rockets themselves, everything appears to be at rest. Since B is halfway between A and C, clearly the two rockets receive the signals at the same time (light travels with the same speed). Imagine, though, that we look at this situation from the point of view of a different observer, outside the rockets and the box, with respect to whom the rockets move with velocity v . This observer sees that rocket A is approaching the light signal, while rocket C is moving away from it (again, remember that light travels with the same speed in each direction). Hence, it is obvious to this observer that rocket A will receive the signal **before** rocket C does.

In the next section we will use the Lorentz equations to prove that:

Events that are simultaneous for one observer and which take place at **different points in space** are not simultaneous for another observer in motion relative to the first.

On the other hand, if two events are simultaneous for one observer and take place at **the same point in space**, they are simultaneous for all other observers as well.

Worked example

A.13 Observer T is in the middle of a train that is moving with constant speed to the right with respect to the train station. Two light signals are emitted at the same time as far as the observer T in the train is concerned (Figure A.22).

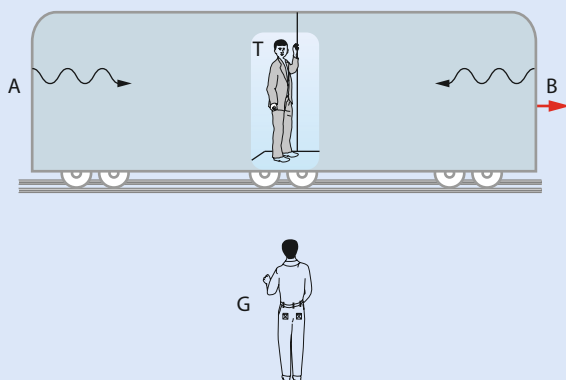


Figure A.22

- a** Determine whether the emissions are simultaneous for an observer G on the ground.
- b** The signals arrive at T at the same time as far as T is concerned.
 - i** Determine whether they arrive at T at the same time as far as G is concerned.
 - ii** Deduce, according to G, which signal is emitted first.

- a** No, because the events (i.e. the emissions from A and from B) take place at different points in space, and so if they are simultaneous for observer T they will not be simultaneous for observer G.
- b i** Yes, because the reception of the two signals by T takes place at the same point in space, so if they are simultaneous for T, they must also be simultaneous for G.
- ii** From G's point of view, T is moving away from the signal from A. So the signal from A has a larger distance to cover to get to T. If the signals are received at the same time, and moved at the same speed c , it must be that the one from A was emitted before that from B.

Simultaneity, like motion, is a relative concept. Our notion of absolute simultaneity is based on the idea of absolute time: events happen at specific times that all observers agree on. Einstein has taught us that the idea of absolute time, just like the idea of absolute motion, must be abandoned.

The Lorentz equations allow for a more quantitative approach to simultaneity. Suppose that as usual we have frames S and S', with frame S' moving to the right with speed v relative to S. Suppose that two events take place at the same time in frame S. The time interval between these two events is thus zero: $\Delta t = 0$. What is the time interval between the same two events when measured in S'? The Lorentz equations give

$$\begin{aligned}\Delta t' &= \gamma \left(\Delta t - \frac{v}{c^2} \Delta x \right) \\ &= -\gamma \frac{v}{c^2} \Delta x\end{aligned}$$

We have the interesting observation that, if $\Delta x \neq 0$, i.e. if the simultaneous events in S occur at the same point in space, then they are also simultaneous in all other frames. However, if $\Delta x \neq 0$, then the events will not be simultaneous in other frames.



Worked examples

A.14 Let us rework the previous example quantitatively. Take the proper length of the train (frame S') to be 300 m and let it move with speed $0.98c$ to the right. Determine, according to an observer on the ground (frame S), which light turns on first and by how much.

We are told that $\Delta t' = 0$ and $\Delta x' = x'_B - x'_A = 300$ m. We want to find Δt , the time interval between the events representing the emission of light from A and from B. The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.98^2}} = 5.0$. Then

$$\begin{aligned}\Delta t &= t_B - t_A \\ &= \gamma \left(\Delta t' + \frac{v}{c^2} \Delta x' \right) \\ &= 5.0 \times \left(0 + \frac{0.98c}{c^2} \times 300 \right) = +4.9 \times 10^{-6} \text{ s}\end{aligned}$$

The positive sign indicates that $t_B > t_A$, that is, the light from A was emitted $4.9 \mu\text{s}$ before the light from B.

Note: if we wanted to find the actual emission times of the lights and not just their difference, we would use

$$t_A = \gamma \left(t' + \frac{v}{c^2} x' \right) = 5.0 \times \left(0 + \frac{0.98c}{c^2} \times (-150) \right) = -2.45 \times 10^{-6} \text{ s}$$

and

$$t_B = \gamma \left(t' + \frac{v}{c^2} x' \right) = 5.0 \times \left(0 + \frac{0.98c}{c^2} \times (+150) \right) = +2.45 \times 10^{-6} \text{ s}$$

giving the same answer for the difference.

A.15 Let us look at the previous example again, but now we will assume that the emissions from A and B happened at the same time for observer G on the ground, in frame S (Figure A.23). Determine which emission takes place first in the train frame, frame S' .

We are now told that $\Delta t = 0$. In frame S , light A is emitted from position $x_A = -a$ and light B from position $x_B = +a$.

We want to find $\Delta t'$. The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.98^2}} = 5.0$. Then

$$\begin{aligned}\Delta t' &= t'_B - t'_A \\ &= \gamma \left(\Delta t - \frac{v}{c^2} \Delta x \right) \\ &= 5.0 \times \left(0 - \frac{0.98c}{c^2} \times 2a \right) = -3.3 \times 10^{-8} a\end{aligned}$$

The negative sign shows that according to observer T, light B was emitted first. But to get a numerical answer we need to know a .

The light signals are emitted from the ends of the train, of proper length 300 m. In the frame S this length is Lorentz-contracted to 60 m, and so $a = 30$ m. (You can also obtain this result as follows. In the frame S , light B is emitted at $t = 0$. The position of this event in S' is $x' = 150$ m.

Using $x' = \gamma(x - vt)$, $150 = 5.0(a - 0) \Rightarrow a = \frac{150}{5.0} = 30$ m.)

Hence, $\Delta t' = -3.3 \times 10^{-8} \times 30 = -9.9 \times 10^{-7}$ s.

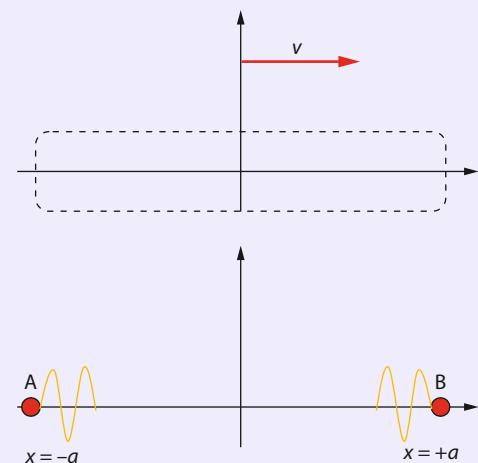


Figure A.23

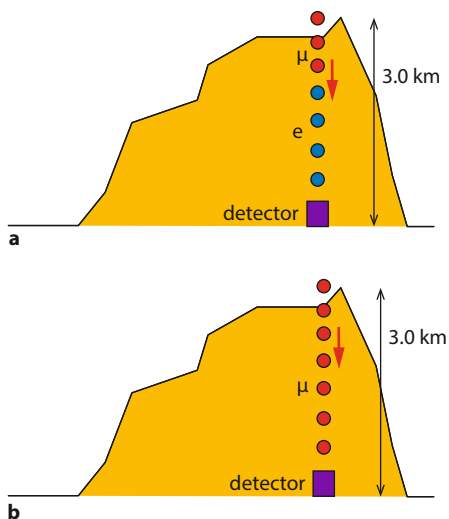


Figure A.24 **a** If relativity did not hold, muons created at the top of the mountain would not have enough time to reach the bottom as muons. **b** Because of relativistic effects, muons do reach the bottom of the mountain.

A2.8 Muon decay

Muons are particles with properties similar to those of electrons except that they are more massive, are unstable and decay to electrons; they have an average lifetime of about 2.2×10^{-6} s. (The reaction is $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$.) This is the lifetime measured in a frame where the muon is at rest: the proper time interval between the creation of a muon and its subsequent decay.

From the point of view of an observer in the laboratory, however, the muons are moving at high speed, and the lifetime is longer because of time dilation. Consider a muon created by a source at the top of a mountain 3.0 km tall (Figure A.24). The muon travels at $0.99c$ towards the surface of the Earth.

Without relativistic time dilation, the muon would have travelled a distance (as measured by ground observers) of only

$$0.99 \times 3 \times 10^3 \times 2.2 \times 10^{-6} \text{ m} = 0.653 \text{ km}$$

before decaying to an electron (Figure A.24a). Thus a detector at the base of the mountain would record the arrival of an electron, not a muon.

But experiments show the arrival of muons at the detector. This is because the lifetime of the muon as measured by ground observers is

$$\begin{aligned} \text{time interval} &= \frac{\text{proper time}}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= \frac{2.2 \times 10^{-6} \text{ s}}{\sqrt{1 - 0.99^2}} \\ &= 1.56 \times 10^{-5} \text{ s} \end{aligned}$$

In this time the muon travels a distance (as measured by ground observers) of $0.99 \times 3 \times 10^3 \times 1.56 \times 10^{-5} \text{ m} = 4.63 \text{ km}$

This means that the muon reaches the surface of the Earth before decaying (Figure A.24b).

The fact that muons do make it to the surface of the Earth is evidence in support of the time dilation effect.

The muon exists as a muon for only 2.2×10^{-6} s in the muon's rest frame. So how does an observer travelling along with the muon explain the arrival of muons (and not electrons) at the surface of the Earth (Figure A.25)?

The answer is that the distance of 3.0 km measured by observers on the Earth is a proper length for them but not for the observer at rest with respect to the muon. This observer claims that it is the Earth that is moving upwards, and so measures a length-contracted distance of

$$3.0 \times \sqrt{1 - 0.99^2} \text{ km} = 0.42 \text{ km}$$

to the surface of the Earth. The Earth's surface is coming up to this observer with a speed of $0.99c$ and so the time when they will meet is

$$\frac{0.42 \times 10^3}{0.99 \times 3 \times 10^8} \text{ s} = 1.4 \times 10^{-6} \text{ s}$$

that is, before the muon decays.



Figure A.25 From the point of view of an observer on the muon, the mountain is much shorter.

In this sense, muon decay experiments are indirect confirmations of the length contraction effect.



Worked example

- A.16** At the Stanford Linear Accelerator, electrons of speed $v = 0.960c$ move a distance of 3.00 km.
- Calculate how long this takes according to observers in the laboratory.
 - Calculate how long this takes according to an observer travelling along with the electrons.
 - Find the speed of the linear accelerator in the rest frame of the electrons.

- a** In the laboratory, the electrons take a time of

$$\frac{3.00 \times 10^3}{0.960 \times 3.00 \times 10^8} \text{ s} = 1.04 \times 10^{-5} \text{ s}$$

- b** The arrival of the electrons at the beginning of the accelerator track and that the end happen at the same point in space as far as the observer travelling with the electrons is concerned, so this is a proper time interval. Thus

$$\text{time interval} = \gamma \times \text{proper time interval}$$

$$\begin{aligned} \gamma &= \frac{1}{\sqrt{1 - 0.960^2}} \\ &= 3.571 \end{aligned}$$

$$1.0420 \times 10^{-5} \text{ s} = 3.571 \times \text{proper time interval}$$

That is,

$$\begin{aligned} \text{proper time interval} &= \frac{1.04 \times 10^{-5}}{3.571} \\ &= 2.91 \times 10^{-6} \text{ s} \end{aligned}$$

- c** The speed of the accelerator is obviously $v = 0.960c$ in the opposite direction. But this can be checked as follows. As far as the electron is concerned, the length of the accelerator track is moving past it and so is length-contracted according to

$$\begin{aligned} \text{length} &= \frac{\text{proper length}}{\gamma} \\ &= \frac{3.00 \text{ km}}{3.571} \\ &= 0.840 \text{ km} \end{aligned}$$

and so has a speed of

$$\begin{aligned} \text{speed} &= \frac{0.840 \times 10^3}{2.91 \times 10^{-6}} \text{ m s}^{-1} \\ &= 2.89 \times 10^8 \text{ m s}^{-1} \\ &\approx 0.96c \end{aligned}$$

Nature of science

Pure science – applying general arguments to special cases

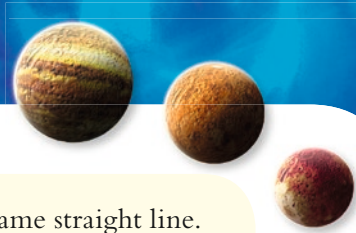
Some years before relativity was introduced by Einstein, an experiment performed by Albert Michelson and Edward Morley gave a very puzzling result concerning the speed of light. The experimenters had expected to find that light travelled faster when it was directed along the path of travel of the Earth than when it was at right angles to the Earth's motion. They found no difference. There were frantic attempts to resolve the difficulties posed by this experiment. One attempt was to assume that moving lengths appear shorter. Lorentz showed that if one used his Lorentz transformation equations, certain difficulties with electromagnetism went away. But it was Einstein who re-derived these transformation equations from far more general principles, the postulates of relativity. By demanding that all inertial observers experience the same laws of physics and measured the same velocity of light in vacuum, the Lorentz equations emerged as the simplest linear equations that could achieve this.

? Test yourself

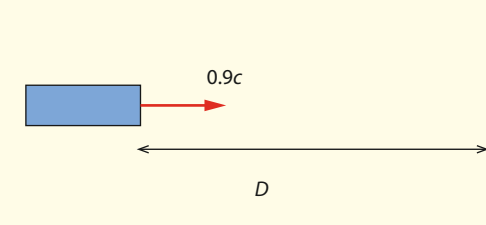
- 10 An earthling sits on a bench in a park eating a sandwich. It takes him 5 min to finish it, according to his watch. He is being monitored by invaders from planet Zenga who are orbiting at a speed of $0.90c$.
 - a Calculate how long the aliens reckon it takes an earthling to eat a sandwich.
 - b The aliens in the spacecraft get hungry and start eating their sandwiches. It takes a Zengan 5 min to eat her sandwich according to Zengan clocks. They are actually being observed by earthlings as they fly over the Earth. Calculate how long it takes a Zengan to eat a sandwich according to Earth clocks.
- 11 A cube has density ρ when the density is measured at rest. Suggest what happens to the density of the cube when it travels past you at a relativistic speed.
- 12 A pendulum in a fast train is found by observers on the train to have a period of 1.0 s. Calculate the period that observers on a station platform would measure as the train moves past them at a speed of $0.95c$.
- 13 A spacecraft moves past you at a speed of $0.95c$ and you measure its length to be 100 m. Calculate the length you would measure if it were at rest with respect to you.
- 14 Two identical fast trains move parallel to each other. An observer on train A tells an observer on train B that by her measurements (i.e. by A's measurements) train A is 30 m long and train B is 28 m long.

The observer on train B takes measurements; calculate what he will find for:

 - a the speed of train A with respect to train B
 - b the length of train A
 - c the length of train B.
- 15 An unstable particle has a lifetime of 5.0×10^{-8} s as measured in its rest frame. The particle is moving in a laboratory with a speed of $0.95c$ with respect to the lab.
 - a Calculate the lifetime of the particle according to an observer at rest in the laboratory.
 - b Calculate the distance travelled by the particle before it decays, according to the observer in the laboratory.
- 16 The star Vega is about 50 ly away from the Earth. A spacecraft moving at $0.995c$ is heading towards Vega.
 - a Calculate how long it will take the spacecraft to get to Vega according to clocks on the Earth.
 - b The crew of the spacecraft consists of 18-year-old IB graduates. Calculate how old the graduates will be (according to their clocks) when they arrive at Vega.



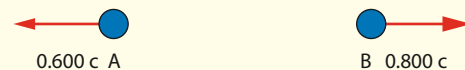
- 17 A rocket travelling at $0.60c$ with respect to the Earth is launched towards a star. After 4.0 yr of travel (as measured by clocks on the rocket) a radio message is sent to the Earth. Calculate when it will arrive on the Earth as measured by:
- observers on the Earth
 - observers on the rocket.
- 18 A spacecraft leaving the Earth with a speed of $0.80c$ sends a radio signal to the Earth as it passes a space station 8.0 ly from the Earth (as measured from the Earth).
- Calculate how long it takes the signal to arrive on the Earth according to Earth observers.
 - Calculate how long it takes the spacecraft to reach the space station (according to clocks on the spacecraft).
 - As soon as the signal is received on the Earth, a reply signal is sent to the spacecraft. Calculate how long the reply signal takes to arrive at the spacecraft, according to Earth clocks.
 - According to spacecraft clocks, calculate how much time goes by between the emission of the signal and the arrival of the reply.
- 19 A rocket approaches a mirror on the ground at a speed of $0.90c$, as shown below. The distance D between the front of the rocket and the mirror is 2.4×10^{12} m, as measured by observers on the ground, when a light signal is sent towards the mirror from the front of the rocket. Calculate when the reflected signal is received by the rocket as measured by:
- observers on the ground
 - observers on the rocket.



- 20 Two objects move along the same straight line. Their speeds are as measured by an observer on the ground. Find:
- the velocity of B as measured by A
 - the velocity of A as measured by B.



- 21 Repeat Question 20 for the arrangement below.



- 22 A particle A moves to the right with a speed of $0.600c$ relative to the ground. A second particle, B, moves to the right with a speed of $0.700c$ relative to A. Calculate the speed of B relative to the ground.
- 23 Particle A moves to the left with a speed of $0.600c$ relative to the ground. A second particle, B, moves to the right with a speed of $0.700c$ relative to A. Find the speed of B relative to the ground.
- 24 A muon travelling at $0.950c$ covers a distance of 2.00 km (as measured by an earthbound observer) before decaying.
- Calculate the muon's lifetime as measured by the earthbound observer.
 - Calculate the lifetime as measured by an observer travelling along with the muon.
- 25 The lifetime of the unstable pion particle is measured to be 2.6×10^{-8} s (when at rest). This particle travels a distance of 20 m in the laboratory just before decaying. Calculate its speed.

In the following questions, the frames S and S' have their usual meanings, that is, S' moves past S with velocity v and, when their origins coincide, both clocks are set to zero.

- 26 In frame S an explosion occurs at position $x = 600$ m and time $t = 2.0 \mu\text{s}$. Frame S' is moving at speed $0.75c$ in the positive direction. Determine where and when the explosion takes place according to frame S' .

- 27 Frame S' moves with speed $0.98c$. An explosion occurs at the origin of frame S' when the clocks in S' read $6.0\ \mu\text{s}$. Calculate where and when the explosion takes place according to frame S .
- 28 Frame S' moves with speed $0.60c$. Calculate the reading of the clocks in frame S' as the origin of S' passes the point $x = 120\text{ m}$.
- 29 Two events in frame S are such that $\Delta x = x_2 - x_1 = 1200\text{ m}$ and $\Delta t = t_2 - t_1 = 6.00\ \mu\text{s}$.
- a i The speed v is $0.600c$. Calculate $\Delta x' = x'_2 - x'_1$ and $\Delta t' = t'_2 - t'_1$.
- ii Determine whether event 1 could cause event 2.
- b Determine whether there is a value of v such that $\Delta t' = 0$. Comment on your answer.
- 30 Two events in frame S are such that $\Delta x = x_2 - x_1 = 1200\text{ m}$ and $\Delta t = t_2 - t_1 = 3.0\ \mu\text{s}$.
- a i The speed v is $0.600c$. Calculate $\Delta x' = x'_2 - x'_1$ and $\Delta t' = t'_2 - t'_1$.
- ii Determine whether event 1 could cause event 2.
- b Determine whether there is a value of v such that $\Delta t' < 0$. Comment on your answer.
- 31 Two simultaneous events in frame S are separated by a distance $\Delta x = x_1 - x_2 = 1200\text{ m}$. Determine the time separating these two events in frame S' , stating which one occurs first. The speed v is $0.600c$.

Learning objectives

- Understand, sketch and work with spacetime diagrams.
- Represent worldlines.
- Explain the twin paradox.

A3 Spacetime diagrams

This section deals with a pictorial representation of relativistic phenomena that makes the concepts of time dilation, length contraction and simultaneity particularly transparent.

A3.1 The invariant hyperbola

The Lorentz transformations have the following important consequence. Consider an event that is measured to have coordinates (x, t) in S and (x', t') in S' . As we know, these coordinates are different in the two frames; the observers disagree about the space and time coordinates. But they all agree on this: the quantity $(x^2 - c^2t^2)$ is the same as $(x'^2 - c^2t'^2)$! We can prove this easily as follows:

$$\begin{aligned}
 (x^2 - c^2t^2) &= \gamma^2(x - vt)^2 - c^2\gamma^2\left(t - \frac{v}{c^2}x\right)^2 \\
 &= \gamma^2\left(x^2 - 2xvt + v^2t^2 - \left(c^2t^2 - 2tvx + \frac{v^2}{c^2}x^2\right)\right) \\
 &= \gamma^2\left(x^2\left(1 - \frac{v^2}{c^2}\right) - (c^2 - v^2)t^2\right) \\
 &= \gamma^2\left(x^2\left(1 - \frac{v^2}{c^2}\right) - c^2\left(1 - \frac{v^2}{c^2}\right)t^2\right) \\
 &= \gamma^2\left(1 - \frac{v^2}{c^2}\right)(x^2 - c^2t^2) \\
 &= x^2 - c^2t^2
 \end{aligned}$$

The usefulness and significance of this result will become apparent in the next section.

A3.2 Spacetime diagrams (Minkowski diagrams)

In Topic 2, we saw lots of motion graphs. In particular, we saw graphs of position (vertical axis) versus time (horizontal axis). In relativity it is customary to show these graphs with the axes reversed, that is, with

Exam tip

Remember that $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$

and so $\gamma^2\left(1 - \frac{v^2}{c^2}\right) = 1$.



time plotted on the vertical axis and position on the horizontal. These are called **spacetime diagrams**. They are also called **Minkowski diagrams** in honour of H. Minkowski, a mathematician friend of Einstein who helped him with the mathematical formulation of the theory. For reasons we will see shortly, it is very convenient to instead plot ct on the vertical axis rather than time itself. The vertical axis then also has units of length. Since the speed of light is a constant everyone agrees on, knowing the value of ct allows us to find the value of t .

Figure A.26 shows a spacetime diagram and an event (marked by the dot). The space and time coordinates of the event are read from the graph in the usual way: we draw lines through the dot parallel to the axes and see where the lines intersect the axes. This event has $x = 5.0$ m and $ct = 6.0$ m, giving $t = \frac{6.0}{3.0 \times 10^8} = 2.0 \times 10^{-8}$ s.

Now consider a particle that is at rest at position $x = a$ (Figure A.27). As time goes by we may think of a sequence of events showing the position of the particle (which does not change) at different times. This sequence of events traces the straight blue line on the spacetime diagram. This is called the **worldline** of the particle.

Another particle that starts from $x = 0$ at $t = 0$ and moves with constant positive velocity would have the worldline shown in red. The speed of this particle is given by

$$v = \frac{\text{distance}}{\text{time}} = \frac{x}{t} = c \frac{x}{ct} = c \tan \theta$$

so $\tan \theta = \frac{v}{c}$.

The tangent of the angle of the worldline with the ct -axis gives the speed, expressed in units of the speed of light.

This is why it is convenient to plot ct on the vertical axis: a photon moves at speed c and so it makes an angle of $\theta = \tan^{-1} \frac{c}{c} = 45^\circ$ with both axes. Since nothing can have a speed that exceeds c , the worldline of any particle will always make an angle less than 45° with the ct -axis. In Figure A.28, the red worldline represents a particle moving to the right. The green worldline belongs to a particle moving to the left. The blue worldline is impossible since it involves a speed greater than c .

Now consider a second inertial reference frame S' that moves to the right with speed v . Let us assume that when clocks in both frames show zero the origins of the frames coincide. The origin of frame S' moves to the right with speed v . Therefore, the worldline of the origin of S' will make an angle $\theta = \tan^{-1} \frac{v}{c}$ with the ct -axis of S . But this worldline is just the collection of all events with $x' = 0$, and is therefore the time axis of the frame S' . Because the speed of light is the same in all frames, the space axis of the new frame must make the same angle with the old x -axis. Thus the new frame has axes as shown in red in Figure A.29. An event will have different coordinates in the two frames, as we know. To find the coordinates of an event in the frame S' , draw lines parallel to the slanted axes and see where they intersect the axes. The coordinates in the two frames are connected by Lorentz transformations.

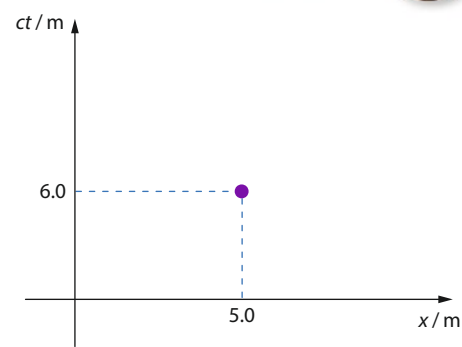


Figure A.26 The space and time coordinates of an event.

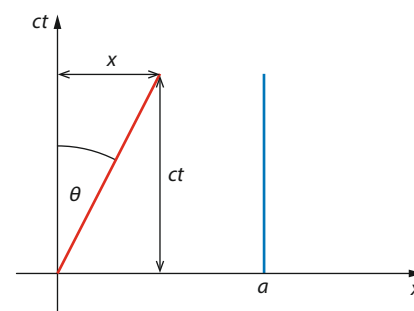


Figure A.27 The worldline of a particle at rest is the sequence of events showing the position of the particle at different times.

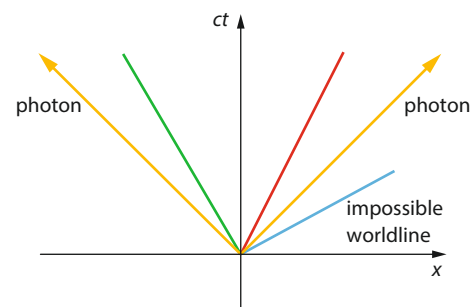


Figure A.28 Various worldlines. The one in blue is impossible because it corresponds to a particle moving faster than light.

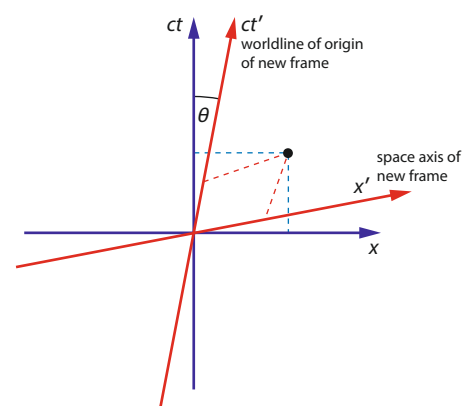


Figure A.29 Two frames with a relative velocity on the same spacetime diagram. The red axes represent a frame moving to the right.

Worked examples

A.17 Use the spacetime diagram below to estimate the speed of frame S' relative to frame S .

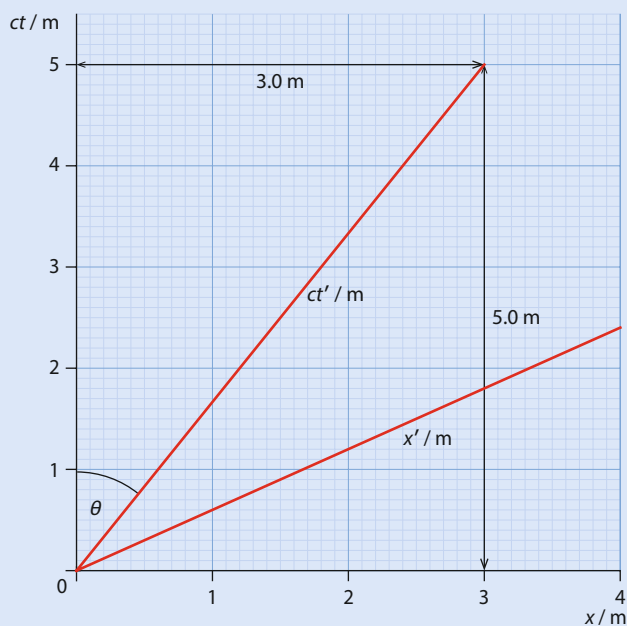


Figure A.30

The tangent of the angle θ is $\tan \theta = \frac{3.0}{5.0} = 0.60$ and so $\frac{v}{c} = 0.60 \Rightarrow v = 0.60c$.

A.18 The spacetime diagram below shows three events. List them from earliest to latest, as observed in each frame.

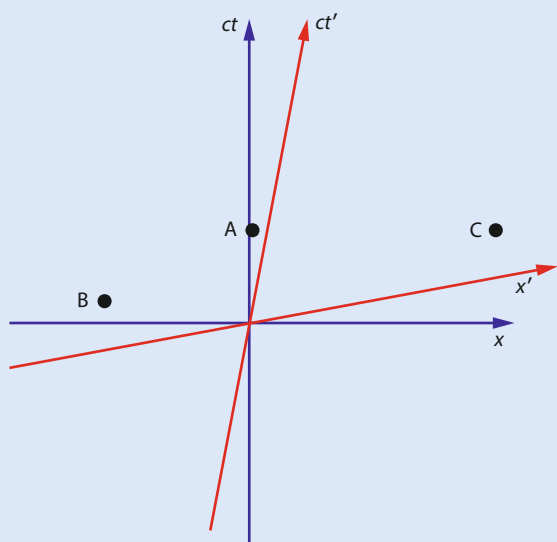


Figure A.31

In S (blue frame), events A and C are simultaneous and occur after event B (A and C are on the same line parallel to the x -axis).

In S' (red frame), events B and C are simultaneous and occur before event A (B and C are on the same line parallel to the x' -axis).



Let us now see the significance of the equation known as **invariant hyperbola**.

Figure A.32 shows the invariant hyperbola passing through a point with coordinates $(x = 1, ct = 0)$. Its equation is $c^2t^2 - x^2 = -1$. All events on the blue curve have the same value of $c^2t^2 - x^2$, namely -1 . Furthermore these events, as observed in frame S' , have the same value of $c^2t'^2 - x'^2$, namely -1 . If $c^2t^2 - x^2$ is plotted on the spacetime axes of frame S , the invariant hyperbola intersects the space axis of frame S' at the point (i.e. event) P . The coordinates of P are $(x', ct' = 0)$. Since $c^2t'^2 - x'^2 = c^2t^2 - x^2$, it follows that $c^2t'^2 - x'^2 = -1$. Since $ct' = 0$, it follows that $x' = 1$ m.

This result shows that the scales on the two space axes (x and x') are not the same.

This is the basis for length contraction, as we will see in the next section.

The scales are also different on the time axes.

A simpler way to see that the scales on the axes are different is to note in Figure A.32 that the blue dashed line passes through the event with coordinates $x = 1, ct = 0$ (in S). It is parallel to the ct' -axis, and intersects the x' -axis at Q . The coordinates of Q are clearly $ct' = 0$ and

$$\begin{aligned} x' &= \gamma(x - vt) \\ &= \gamma(1 - 0) \\ &= \gamma \end{aligned}$$

This sets the scale on the x' -axis.

Consider now the invariant hyperbola that passes through the point with coordinates $(ct = 1, x = 0)$. Its equation is $c^2t^2 - x^2 = 1 - 0 = 1$. This is plotted in Figure A.33.

The hyperbola intersects the time axis of frame S' at event P , whose coordinates are $(ct', 0)$. Since $(ct')^2 - (x')^2 = c^2t^2 - x^2$, it follows that $(ct')^2 - (x')^2 = 1$, and therefore at event P , $(ct')^2 - 0 = 1$ and so $ct' = 1$ m. We again see that the scales on the two time axes are different. This is the basis for time dilation.

Again, a simpler way to see the difference in scale is to note in Figure A.33 that the blue dashed line passes through the event with coordinates $x = 0, ct = 1$ m (in S). It intersects the ct' -axis at Q . Its coordinates are $x' = 0$ and

$$\begin{aligned} t' &= \gamma\left(t - \frac{vx}{c^2}\right) \\ ct' &= \gamma\left(ct - \frac{vx}{c}\right) \\ ct' &= \gamma(1 - 0) \\ ct' &= \gamma \end{aligned}$$

This sets the scale on the ct' -axis.

Exam tip

The mathematical details of the invariant hyperbola are not required in examinations. However, you should understand why the scales on the axes of frame S and frame S' are different. The important thing to remember is that the scales are in fact different.

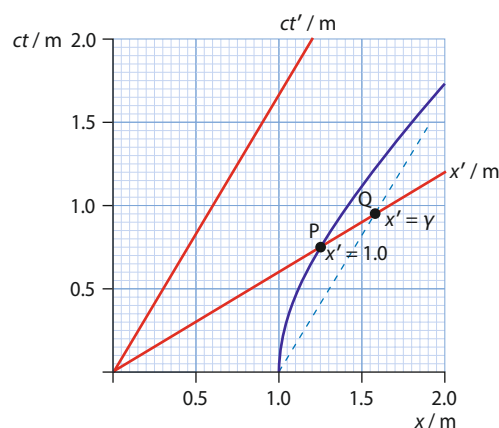


Figure A.32 a The invariant hyperbola sets the scale on the x' -axis (red frame). **b** This scale is also set by the dashed blue line, which is parallel to the ct' -axis and passes through $x = 1$ m.

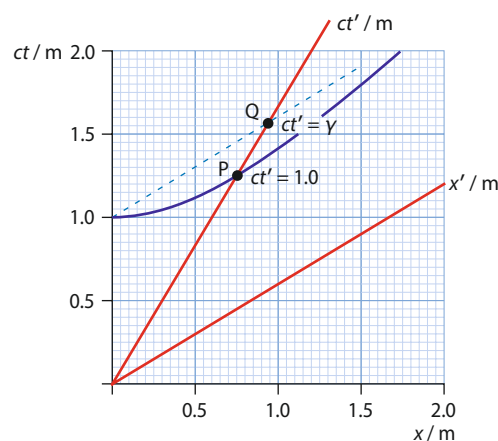


Figure A.33 The invariant hyperbola sets the scale on the ct' -axis (red frame).

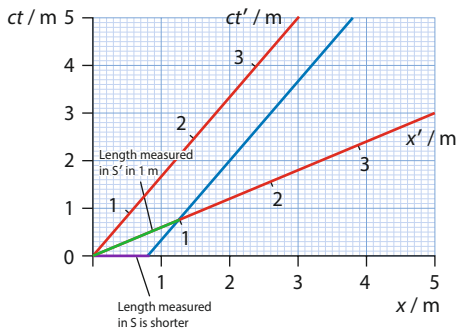


Figure A.34 Rod at rest in S' . Observers in S' measure a length of 1 m for the rod. Observers in S measure a shorter length.

Exam tip

For observers in S , the rod is moving. So its length must be measured when the ends are recorded at the same time.

Exam tip

Do not be led astray by your knowledge of Euclidean geometry! The length in S' 'looks' longer, but it isn't. The scales on the two axes are not the same.

A3.3 Length contraction and spacetime diagrams

Imagine a rod of length 1 m at rest in the frame S' . At $t' = 0$ the left end of the rod is at $x' = 0$ and the other end is at $x' = 1$ m. The worldlines of the ends of the rod are shown in Figure A.34. The left end of the rod has a worldline that is the ct' -axis; the worldline of the other end is the blue line (which is parallel to the ct' -axis).

The worldlines of the two ends intersect the x -axis at 0 and 0.8 m. These two intersection points represent the positions of the moving rod's ends **at the same time** in frame S . Their difference therefore gives the length of the rod as measured in S . The length is 0.8 m, less than 1 m, the length in S' . The rod which is moving according to observers in S has contracted in length when measured in frame S .

What if we had a rod of proper length 1 m at rest in the frame S which was viewed by observers in the frame S' ? This is illustrated in Figure A.35.

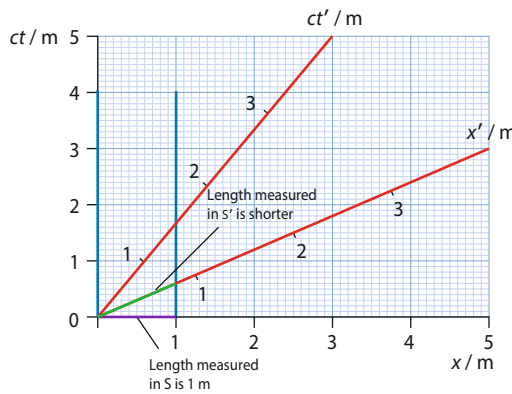


Figure A.35 Rod at rest in S . Observers in S measure a length of 1 m for the rod. The blue lines are the worldlines of the ends of the rod. Observers in S' measure a shorter length.

The two thick blue lines represent the worldlines of the ends of the rod. They intersect the x' -axis at 0 and 0.8 m. These two intersection points represent the positions of the rod's ends **at the same time** in frame S' . Their difference therefore gives the length of the rod as measured in S' . The length is 0.8 m, less than 1 m, the length in S . The rod which is moving according to observers in S' has contracted in length when measured in frame S' .

A3.4 Time dilation and spacetime diagrams

Figure A.36 is a spacetime diagram showing the standard frames S and S' . A clock at rest at the origin of S' ticks at O and then at P . The time between ticks is $ct' = 1$ m. What is the time between ticks according to frame S ? We have to draw a line parallel to the x -axis through P . This intersects the S time axis at point Q . This interval is greater than 1. Since P has coordinates $(ct' = 1 \text{ m}, x' = 0)$ in S' , this event in the frame S has time coordinate

$$t = \gamma \left(t' + \frac{vx'}{c^2} \right) = \gamma t'$$

$$ct = \gamma ct'$$

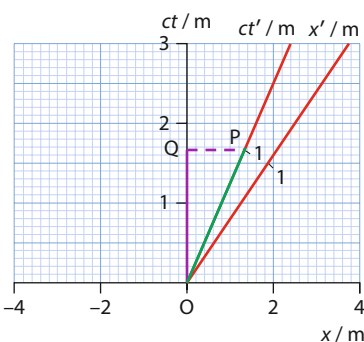


Figure A.36 The green line is the worldline of a clock at rest in frame S' . This clock shows $ct' = 1$ at P .



Similarly, Figure A.37 can be used to analyse a clock at rest at the origin of frame S. It ticks at O and then at P. The time between ticks is such that $ct=1$. From P we draw a line parallel to the x' -axis, which intersects the ct' -axis at Q. The interval OQ is longer than 1 m.

A3.5 The twin paradox

We have seen that the effect of time dilation is symmetric: if you move relative to me, I say that your clocks are running slow compared with mine but you say that my clocks are running slow compared with yours. This symmetry of time dilation is shown clearly in the last two figures in the previous section. An issue arises when two clocks, initially at the same place and showing the same time (say zero), are separated. Suppose one clock moves at a relativistic speed away, suddenly reverses direction and comes back to its starting place next to the clock that stayed behind. The readings of the two clocks are compared. What do they show? In the **twin paradox** version of the story, the clocks are replaced by twins. Jane stays on the Earth and claims that since her twin brother Joe is the one who moved away he must be younger. But Joe can claim that it is Jane and the Earth that moved away and then came back, so Jane should be younger. Who is the younger of the two when the twins are reunited?



Deeply ingrained ideas are hard to get rid of

The physicist W. Rindler, discussing the twin paradox in his book *Introduction to Special Relativity* (Oxford University Press, 1991), says:

It is quite easily resolved, but seems to possess some hidden emotional content that makes it the subject of interminable debate among the dilettantes of relativity.

So deeply ingrained is the idea of absolute time in us that we find this example hard to accept.

The resolution of the ‘paradox’ is that Jane has been in the same inertial frame throughout, whereas Joe changed inertial frames in the interval of time it took for him to turn around. He was in an inertial frame moving outwards but he changed to one moving inwards for the return trip. During the changeover from one frame to the other he must have experienced acceleration and forces which Jane never did. The acceleration makes this situation asymmetric, and Joe is the one who is younger on his return to Earth.

To understand this quantitatively, suppose Joe leaves the Earth in a rocket at a speed of $0.6c$, travels to a distant planet 3.0ly away (as measured by Earth observers) and returns. The gamma factor γ is 1.25. As Jane sees it, his outbound trip takes $\frac{3.0\text{ly}}{0.6c} = 5$ yr, and another 5 yr to return. She will have aged 10 years when her brother returns. For Joe, time is running slower by the gamma factor, so he will age by $\frac{5.0}{1.25} = 4$ yr on the outward trip and another 4 yr on the way back. He will be 8 years older. We can show all of this on a space–time diagram (Figure A.38).

Exam tip

Do not be led astray by your knowledge of Euclidean geometry! In triangle OPQ the hypotenuse is the side OP and so it would seem to be the longest side. But the geometry of spacetime diagrams is not Euclidean. The Pythagorean theorem does not hold.

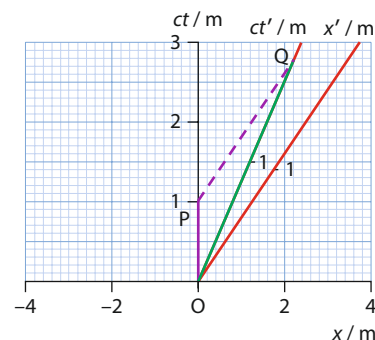


Figure A.37 The purple line is the worldline of a clock at rest in frame S. This clock shows $ct=1$ at P.

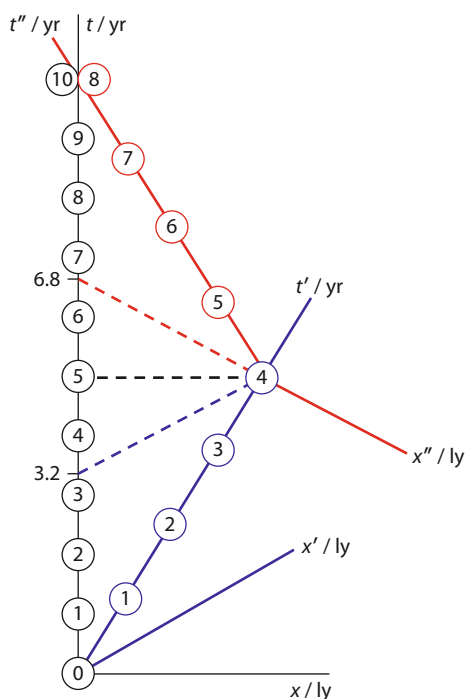


Figure A.38 Spacetime diagram used to resolve the twin paradox. (Adapted from N. David Mermin, *It's About Time*, Princeton University Press, 2005)

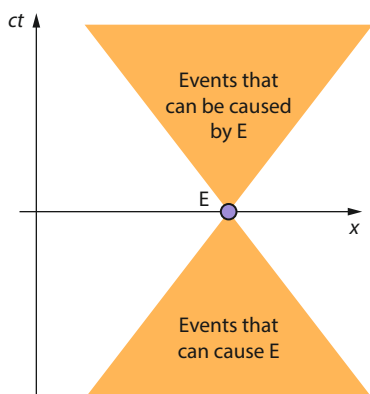


Figure A.39 The past and future light cones of event E.

The black frame, S , is Jane's. The blue frame, S' , is Joe's on the way out, and the red frame, S'' , is Joe's on the way back. The circles show what the clocks read in each frame. We see that, at the moment before Joe turns around and changes frame, the Earth clock shows 5 yr and Joe's clock shows 4 yr. When Joe's clock shows 4 yr, he determines (by drawing a line parallel to the x' -axis) that Earth clocks show 3.2 yr. (Time dilation is symmetric!) As soon as Joe changes frame, he determines (by drawing a line parallel to the x'' -axis) that Earth clocks show 6.8 yr. The sudden switch from one frame to another seems to create a missing time of $10 - 2 \times 3.2 = 3.6$ yr. But this is not mysterious and nothing strange has happened to either Jane or Joe during this time; it has to do with the fact that we have various frames. The 'now' of frame S' is found by drawing lines parallel to the x' -axis. The 'now' of frame S'' is found by drawing lines parallel to the x'' -axis. Thus, what the two frames understand by 'now' is different.

A3.6 Causality

Figure A.39 shows an event E on a spacetime diagram. The cones above and below E are defined respectively by the timelines of photons leaving E and arriving at E. The cone formed by the arriving photons (the past light cone) includes all events that could in principle have caused E to occur. Since nothing can move faster than light, no event in the past of E outside the shaded light cone could have influenced E. Similarly, the future light cone of E consists of those events that could in principle be caused by E. E cannot influence any event outside this light cone.

Nature of science

The power of diagrams – again

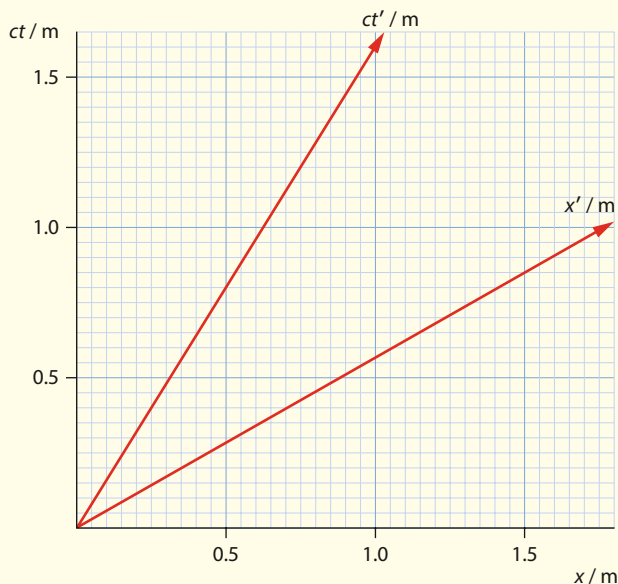
Spacetime diagrams offer a simple and pictorial way of understanding relativity. They can be used to unambiguously resolve misunderstandings and 'paradoxes'. Their power lies in their simplicity and their clarity. In physics, using diagrams with appropriate notation to describe situations has always helped understanding. Spacetime diagrams are especially useful in showing what event can or cannot cause another. Feynman diagrams are another example of visualisation of a model, as is the use of vectors to show magnitude and direction.



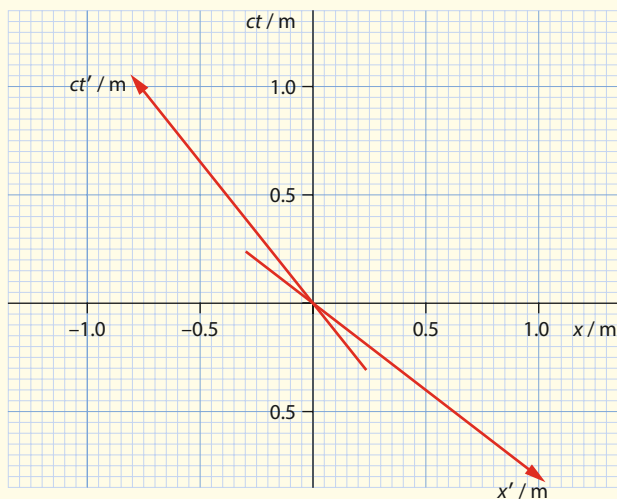
? Test yourself

In the questions that follow, the spacetime diagrams represent two inertial frames. The black axes represent frame S. The red axes represent a frame S' that moves past frame S with velocity v .

32 Use the spacetime diagram to calculate the velocity of frame S' relative to S.

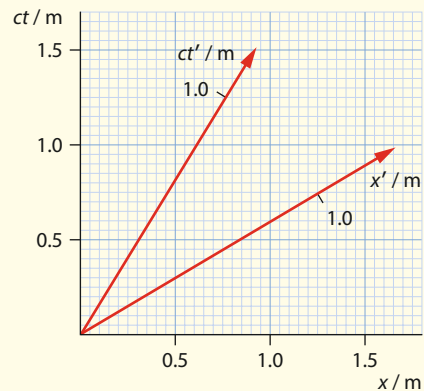


33 Use the space–time diagram to calculate the velocity of frame S' relative to S.



34 a On a copy of the spacetime diagram, draw the worldline of a particle that is:

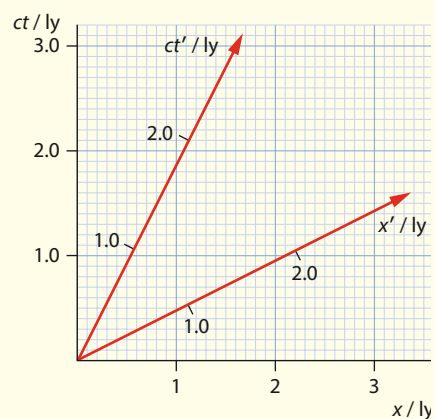
- i at rest in frame S at $x = 1.0\text{ m}$
- ii moving with velocity $-0.80c$ as measured in S at $x = 1.0\text{ m}$ at $t = 0$.



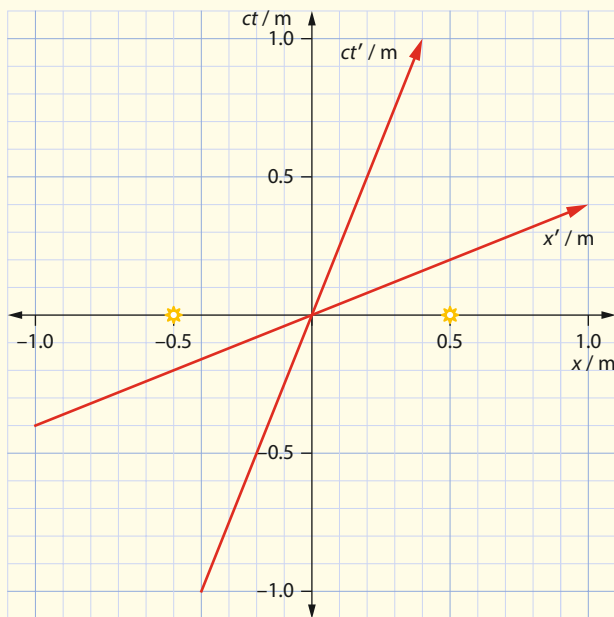
b A photon is emitted from position $x = +1.0\text{ m}$ at $t = 0$. Use the spacetime diagram from part a to estimate when the photon arrives at the eye of an observer at rest at the origin of frame S', as observed i in S and ii in S'.

35 a S' represents an alien attack cruise ship. Use the diagram below to determine the speed of the cruise ship relative to S.

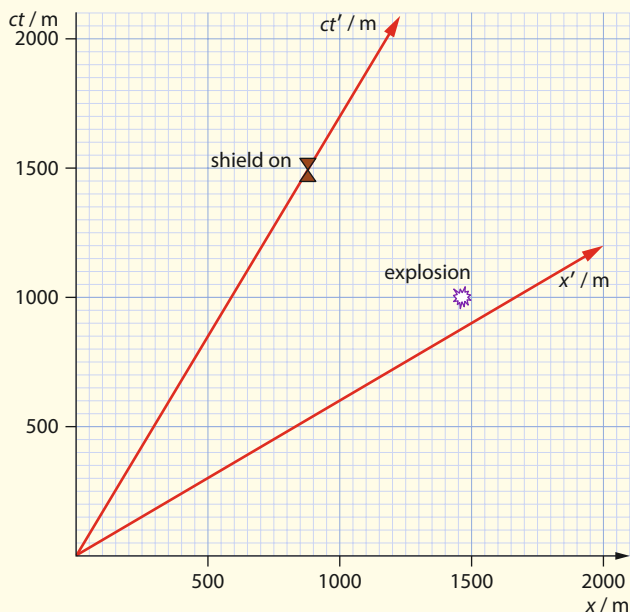
b At $t = 0$, a laser beam moving at the speed of light is launched from $x = +3.0\text{ ly}$ towards the cruise ship. By drawing the worldline of the laser beam, estimate the time when the beam hits the cruise ship, as observed i in S and ii in S'.



- 36 The diagram shows two lamps at $x = \pm 0.5$ m that turn on at $t = 0$ in frame S.

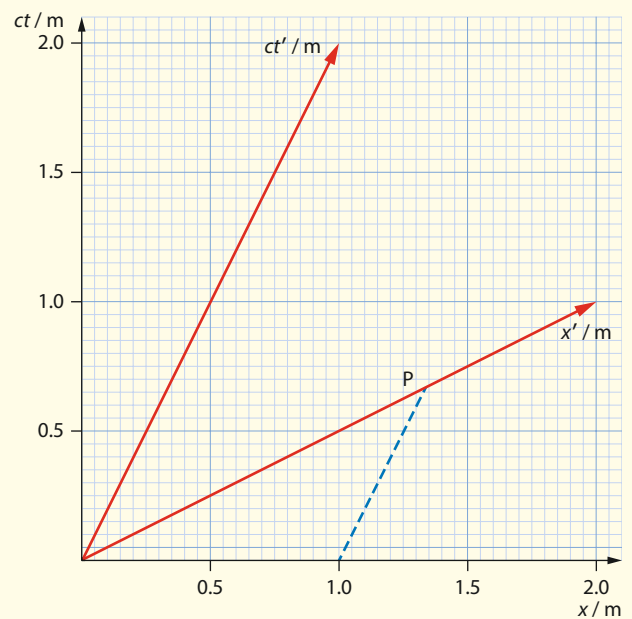


- Determine which lamp turns on first as observed in frame S' .
 - Draw the worldlines of the photons emitted from the two lamps towards an observer at rest at the origin of S' .
 - Identify the lamp whose light reaches an observer at the origin of frame S' first.
- 37 In the spacetime diagram below, the red axes represent a spacecraft that is moving past a planet (black axes). An explosion on the planet takes place at $x = 1500$ m and $ct = 1000$ m, emitting deadly photons. A shield around the spacecraft is turned on at the indicated point in spacetime.

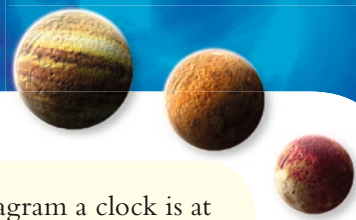


- Determine if the spacecraft will be saved.
- Using Lorentz transformations or otherwise, calculate, in the spacecraft frame:
 - the time of the explosion
 - the position of the explosion.

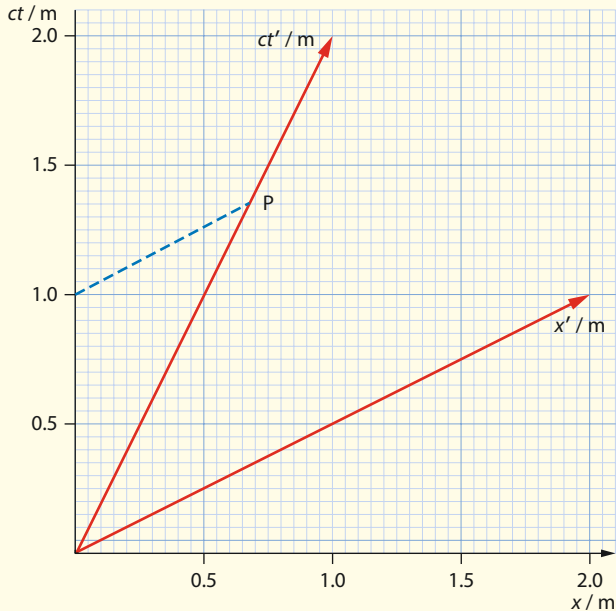
- 38 a The dashed blue line in the spacetime diagram below is parallel to the ct' -axis and intersects the x' -axis at P. Using Lorentz transformations, find the coordinates of P in the frame S' , and hence label the event with coordinates $(x' = 1 \text{ m}, ct' = 0)$.



- Repeat part a, where now the speed v is arbitrary. Express your answer in terms of the gamma factor, γ .

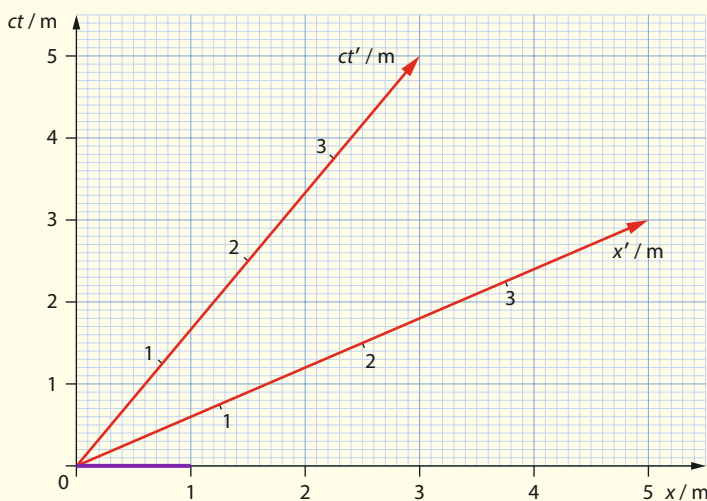


- 39 a The dashed blue line in the spacetime diagram is parallel to the x' -axis and intersects the ct' -axis at P. Find the coordinates of P in the frame S' , and hence label the event with coordinates $(x'=0, ct'=1 \text{ m})$.



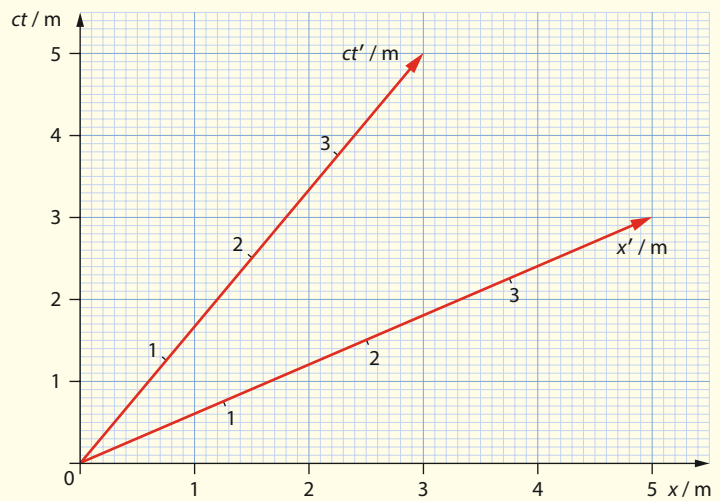
- b Repeat part a, where now the speed v is arbitrary. Express your answer in terms of the gamma factor γ .

- 40 The purple line in the spacetime diagram below represents a rod of length 1.0 m at $t=0$ as measured in the frame S .



- a On a copy of this diagram, draw appropriate lines to show that the length of the rod measured in frame S' is less than 1 m.
- b Draw a line to represent a rod of length 1.0 m as measured in the frame S' . By drawing appropriate lines, show that the length of the rod measured in frame S is less than 1 m.

- 41 In the following spacetime diagram a clock is at rest at the origin of frame S' .



Its first tick occurs at $t'=0$; mark this event on the diagram. The second tick of the clock occurs when $ct' = 1 \text{ m}$ as measured in S' ; mark this event on the same spacetime diagram. By drawing appropriate lines, estimate the time in between ticks as measured in S .

Learning objectives

- Understand and use the concepts of total and rest energy.
- Work with **relativistic momentum**.
- Solve problems with particle acceleration.
- Appreciate the invariance of electric charge.
- Appreciate photons as massless relativistic particles.
- Work with relativistic units.

Exam tip

The phrase ‘to create a particle from the vacuum’ refers to particle collisions in particle accelerators in which energy, usually in the form of photons, materialises as particle–antiparticle pairs.

A4 Relativistic mechanics (HL)

This section introduces the changes in Newtonian mechanics that are necessary as a result of the postulates of relativity. We will have to modify the concepts of **total energy** and momentum.

A4.1 Relativistic energy and rest energy

One of the first consequences of Einstein’s theories in mechanics is the equivalence of mass and energy. The theory of relativity predicts that, to a particle of mass m that is at rest with respect to some inertial observer, there corresponds an amount of energy E_0 that the observer measures to be $E_0 = mc^2$. (We will not be able to give a proof of this statement in this book.) This energy is called the **rest energy** of the particle.

Rest energy is the amount of energy needed to produce a particle at rest.

Similarly, if the particle moves with speed v relative to some inertial observer, the energy corresponding to the mass of the particle that this observer will measure is given by

$$E = \gamma mc^2 \\ = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

This is energy that the particle has because it has mass and because it moves. If the particle has other forms of energy, such as gravitational potential energy or electrical potential energy, then the total energy of the particle will be γmc^2 plus the other forms of energy. We will mostly deal with situations where the particle has no other forms of energy associated with it. The total energy of the particle in that case is then just γmc^2 .

It is important to note immediately that, as the speed of the particle approaches the speed of light, the total energy approaches infinity (Figure A.40). This is a sign that a particle with mass cannot reach the speed of light. Only particles without mass, such as photons, can move at the speed of light.

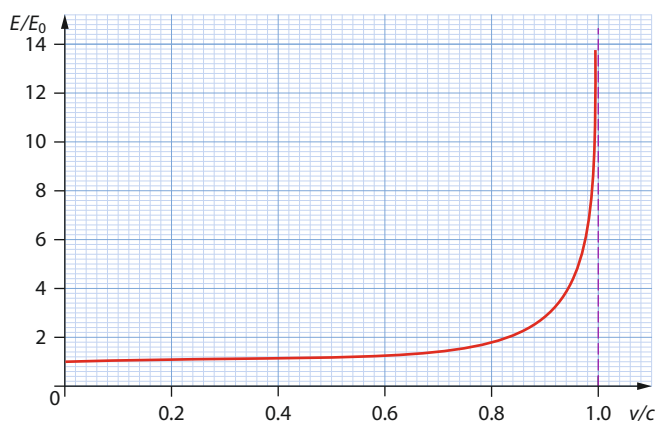


Figure A.40 A graph of the variation with $\frac{v}{c}$ of the ratio of the total energy E to the rest energy E_0 of a particle. As the speed of the particle approaches the speed of light, its energy increases without limit.



Worked examples

A.19 Find the speed of a particle whose total energy is double its rest energy.

We have that

$$E = \gamma mc^2$$

$$2mc^2 = \gamma mc^2$$

$$\Rightarrow \gamma = 2$$

$$\Rightarrow \sqrt{1 - \frac{v^2}{c^2}} = \frac{1}{2}$$

$$\Rightarrow 1 - \frac{v^2}{c^2} = \frac{1}{4}$$

$$\Rightarrow \frac{v^2}{c^2} = \frac{3}{4}$$

$$\Rightarrow v = \frac{\sqrt{3}}{2}c = 0.866c$$

A.20 Find **a** the rest energy of an electron and **b** its total energy when it moves at a speed equal to $0.800c$.

a The rest energy is

$$\begin{aligned} E &= mc^2 \\ &= 9.1 \times 10^{-31} \times 9 \times 10^{16} \text{ J} \\ &= 8.19 \times 10^{-14} \text{ J} \\ &= \frac{8.19 \times 10^{-14}}{1.6 \times 10^{-19}} \text{ eV} \\ &= 0.511 \text{ MeV} \end{aligned}$$

b The gamma factor at a speed of $0.80c$ is

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - 0.8^2}} = \frac{1}{0.6} = 1.667$$

and so the total energy is

$$E = \gamma mc^2 = 1.667 \times 0.511 \text{ MeV} = 0.852 \text{ MeV}$$

A.21 A proton (rest energy 938 MeV) has a total energy of 1170 MeV. Find its speed.

Since the total energy is given by $E = \gamma mc^2$, we have that

$$1170 = \frac{938}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$\sqrt{1 - \frac{v^2}{c^2}} = 0.8017$$

$$1 - \frac{v^2}{c^2} = 0.6427$$

$$\frac{v^2}{c^2} = 0.3573$$

$$v = 0.598c$$

If a particle is accelerated through a potential difference of V volts, its total energy will increase by an amount qV , where q is the charge of the particle. If the particle is initially at rest, its initial total energy is the rest energy, $E_0 = mc^2$. After going through the potential difference, the total energy will be $E = mc^2 + qV$. We can then find the speed of the particle, as the next example shows.

Worked examples

A.22 An electron of rest energy 0.511 MeV is accelerated through a potential difference of 5.0 MV in a lab.

- Find its total energy with respect to the lab.
- Find its speed with respect to the lab.

a The total energy of the electron will increase by

$$qV = 1e \times 5.0 \times 10^6 \text{ volts} = 5.0 \text{ MeV}$$

and so the total energy is

$$\begin{aligned} E &= m_0c^2 + qV = 0.511 \text{ MeV} + 5.0 \text{ MeV} \\ &= 5.511 \text{ MeV} \end{aligned}$$

b We know that

$$E = \gamma mc^2$$

$$5.511 = \gamma \times 0.511$$

$$\gamma = \frac{5.511}{0.511}$$

$$= 10.785$$

Since $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$, it follows that

$$10.785 = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$\sqrt{1 - \frac{v^2}{c^2}} = \frac{1}{10.785} (= 0.0927)$$

$$1 - \frac{v^2}{c^2} = 0.008597$$

$$v = 0.996c$$



A.23 a A proton is accelerated from rest through a potential difference V . Calculate the value of V that will cause the proton to accelerate to a speed of $0.95c$. (The rest energy of a proton is 938 MeV .)

b Determine the accelerating potential required to accelerate a proton from a speed of $0.95c$ to a speed of $0.99c$.

a The gamma factor at a speed of $0.95c$ is

$$\begin{aligned}\gamma &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= \frac{1}{\sqrt{1 - 0.95^2}} \\ &= 3.20\end{aligned}$$

The total energy of the proton after acceleration is thus

$$\begin{aligned}E &= \gamma mc^2 \\ &= 3.20 \times 938\text{ MeV} \\ &= 3002\text{ MeV}\end{aligned}$$

From

$$E = mc^2 + qV$$

we find

$$qV = (3002 - 938)\text{ MeV} = 2064\text{ MeV}$$

and so

$$V = 2.1 \times 10^9\text{ V}$$

Notice how we have avoided using SI units in order to make the numerical calculations easier.

b The total energy of the proton at a speed of $0.95c$ is (from part **a**) $E = 3002\text{ MeV}$. The total energy at a speed of $0.99c$ is (working as in **a**) $E = 6649\text{ MeV}$. The extra energy needed is then $6649 - 3002 = 3647\text{ MeV}$, so the accelerating potential must be $3.6 \times 10^9\text{ V}$. Notice that a larger potential difference is needed to accelerate the proton from $0.95c$ to $0.99c$ than from rest to $0.95c$. This is a sign that it is impossible to reach the speed of light. (See also the next example.)

- A.24** A constant force is applied to a particle which is initially at rest. Sketch a graph that shows the variation of the speed of the particle with time for
- Newtonian mechanics
 - relativistic mechanics.

In Newtonian mechanics, a constant force produces a constant acceleration, and so the speed increases uniformly without limit, exceeding the speed of light. In relativistic mechanics, the speed increases uniformly as long as the speed is substantially less than the speed of light, and is essentially identical with the Newtonian graph. However, as the speed increases, so does the energy. Because it takes an infinite amount of energy for the particle to reach the speed of light, we conclude that the particle never reaches the speed of light. The speed approaches the speed of light asymptotically. Note that the speed is always less than the corresponding Newtonian value at the same time. Hence we have the graph shown in Figure A.41.

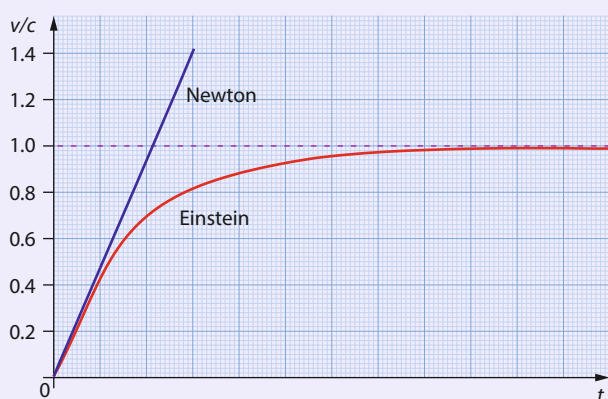


Figure A.41

A4.2 Momentum and energy

In the last section we saw the need to change the definition of total energy. The new definition ensures that a particle cannot be accelerated to the speed of light, because that would take an infinite amount of energy. One other change is required, to momentum, so that in relativistic collisions momentum is conserved as it is in ordinary mechanics.

In classical mechanics, momentum is given by the product of mass and velocity, but in relativity this is modified to

$$p = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$= \gamma mv$$

We still have the usual law of momentum conservation, which states that when no external forces act on a system the total momentum stays the same. The symbol m here stands for the rest mass of the particle and is a constant for all observers.

Unlike Newtonian mechanics, a **constant** force on the particle will produce a **decreasing** acceleration in such a way that the speed never reaches the speed of light.



Worked example

A.25 A constant force F acts on an electron that is initially at rest. Find the speed of the electron as a function of time.

Initially, for small t , the speed increases uniformly, as in Newtonian mechanics. But as t becomes large, the speed tends to the speed of light, but does not reach or exceed it. This is because as the speed increases the acceleration becomes smaller and smaller, and the speed never reaches the speed of light. This results in the graph shown in Figure A.42.

Begin with Newton's second law, $F = \frac{dp}{dt}$, where $p = \gamma mv$ is the momentum of the electron. Since

$$\begin{aligned} \frac{dp}{dt} &= \frac{d}{dt} \left(\frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}} \right) \\ &= \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{dv}{dt} + \left(\frac{v}{c^2} \right) \frac{m}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} \frac{dv}{dt} \end{aligned}$$

$$\frac{dp}{dt} = \frac{m}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} \frac{dv}{dt}$$

it follows that

$$\frac{dv}{dt} = \frac{F}{m} \left(1 - \frac{v^2}{c^2}\right)^{3/2}$$

or

$$\frac{dv}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} = \frac{F}{m} dt$$

Integrating both sides, we find that (assuming the mass starts from rest)

$$\frac{v}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{F}{m} t$$

Solving for the speed, we find

$$v^2 = \frac{(Ft)^2}{(mc)^2 + (Ft)^2 c^2}$$

The way the speed approaches the speed of light is shown in Figure A.42.

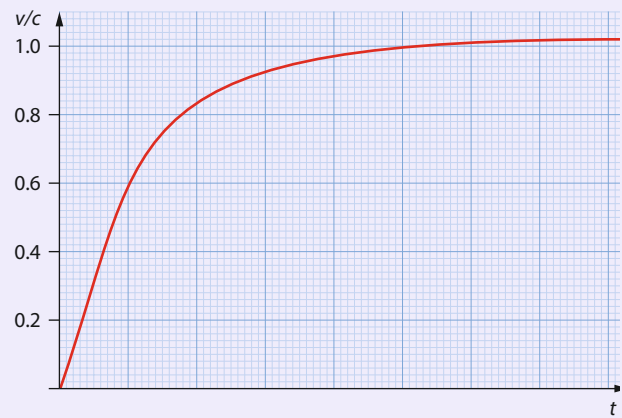


Figure A.42

A4.3 Kinetic energy

A mass moving with velocity v has a total energy E given by

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Its kinetic energy E_k is defined as the total energy minus the rest energy:

$$E_k = E - mc^2$$



Newtonian mechanics is a good approximation to relativity at low speeds

This relativistic definition of kinetic energy does not look similar to the ordinary kinetic energy, $\frac{1}{2}mv^2$. In fact, when v is small compared with c , we can approximate the value of the relativistic

factor $\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$ using the binomial expansion for $\frac{1}{\sqrt{1-x}}$ for small x :

$$\frac{1}{\sqrt{1-x}} \approx 1 + \frac{1}{2}x + \dots$$

Applying this to E_k with $x = \left(\frac{v}{c}\right)^2$, we find

$$E_k = mc^2 \left(1 + \frac{1}{2}\frac{v^2}{c^2} + \dots\right) - mc^2$$

That is,

$$E_k \approx \frac{1}{2}mv^2$$

In other words, for low speeds the relativistic formula reduces to the familiar Newtonian version. For higher speeds, the relativistic formula must be used.

This can be rewritten as

$$\begin{aligned} E_k &= \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} - mc^2 \\ &= \gamma mc^2 - mc^2 \\ &= (\gamma - 1)mc^2 \end{aligned}$$

This definition ensures that the kinetic energy is zero when $v=0$, as can easily be checked. As in ordinary mechanics, the work done by a net force equals the change in kinetic energy in relativity as well.



Total energy, momentum and mass are related: from the definition of momentum, we find that

$$\begin{aligned}
 p^2 c^2 + m^2 c^4 &= \frac{m^2 v^2 c^2}{1 - \frac{v^2}{c^2}} + m^2 c^4 \\
 &= \frac{m^2 c^4}{1 - \frac{v^2}{c^2}} \\
 &= E^2
 \end{aligned}$$

That is,

$$E = \sqrt{m^2 c^4 + p^2 c^2}$$

(This formula is the relativistic version of the conventional formula

$E = \frac{p^2}{2m}$ from Newtonian mechanics.) This relation can be remembered

by using the Pythagorean theorem in the triangle in Figure A.43.

The formula we derived above applies also to those particles that have zero mass, such as the photon, in which case $E = pc$. Remembering that, for a photon, $E = hf$, we have $p = \frac{hf}{c} = \frac{h}{\lambda}$.

A4.4 Relativistic units

We can use units such as $\text{MeV } c^{-2}$ or $\text{GeV } c^{-2}$ for mass. This follows from the fact that the rest energy of a particle is given by $E_0 = mc^2$ and so allows us to express the mass of the particle in terms of its rest energy as $m = E_0/c^2$. Thus, the statement 'the mass of the pion is $135 \text{ MeV } c^{-2}$ ', means that the rest energy of this particle is $135 \text{ MeV } c^{-2} \times c^2 = 135 \text{ MeV}$. (To find the mass in kilograms, we would first have to convert MeV to joules and then divide the result by the square of the speed of light.)

Similarly, the momentum of a particle can be expressed in units of $\text{MeV } c^{-1}$ or $\text{GeV } c^{-1}$. A particle of rest mass $5.0 \text{ MeV } c^{-2}$ and total energy 13 MeV has a momentum given by

$$\begin{aligned}
 E^2 &= m^2 c^4 + p^2 c^2 \\
 \Rightarrow p^2 c^2 &= (169 - 25) \text{ MeV}^2 \\
 &= 144 \text{ MeV}^2 \\
 \Rightarrow pc &= 12 \text{ MeV} \\
 \Rightarrow p &= 12 \text{ MeV } c^{-1}
 \end{aligned}$$

Exam tip

In the last section we saw, for particle acceleration through a potential difference V , that $qV = \Delta E$. But ΔE is also ΔE_k since, for a particle accelerated from rest, $\Delta E = \gamma mc^2 - mc^2 = (\gamma - 1)mc^2 = \Delta E_k$

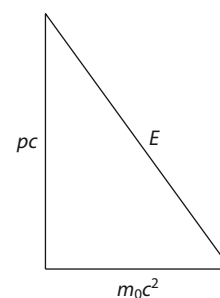


Figure A.43 The rest energy, momentum and total energy are related through the Pythagorean theorem for the triangle shown.

Exam tip

It is absolutely essential that you are comfortable with these units.

The speed can be found from

$$\begin{aligned} E &= \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \\ \Rightarrow 13 &= \frac{5}{\sqrt{1 - \frac{v^2}{c^2}}} \\ \Rightarrow \sqrt{1 - \frac{v^2}{c^2}} &= \frac{5}{13} \\ \Rightarrow \frac{v^2}{c^2} &= \frac{144}{169} \\ \Rightarrow v &= \frac{12}{13}c \end{aligned}$$

In conventional SI units, a momentum of $12 \text{ MeV } c^{-1}$ is

$$\begin{aligned} &12 \times 10^6 \times 1.6 \times 10^{-19} \frac{\text{J}}{3 \times 10^8 \text{ ms}^{-1}} \\ &= 6.4 \times 10^{-21} \frac{\text{kgm}^2 \text{ s}^{-2}}{\text{ms}^{-1}} \\ &= 6.4 \times 10^{-21} \text{ kgms}^{-1} \end{aligned}$$

Worked examples

A.26 Find the momentum of a pion (rest mass $135 \text{ MeV } c^{-2}$) whose speed is $0.80c$.

The total energy is

$$\begin{aligned} E &= \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \\ &= \frac{135}{\sqrt{1 - 0.80^2}} \\ &= 225 \text{ MeV} \end{aligned}$$

Using

$$E^2 = m^2c^4 + p^2c^2$$

We obtain

$$\begin{aligned} pc &= \sqrt{225^2 - 135^2} \\ &= 180 \text{ MeV} \\ \Rightarrow p &= 180 \text{ MeV } c^{-1} \end{aligned}$$



A.27 Find the speed of a muon (rest mass = $105 \text{ MeV } c^{-2}$) whose momentum is $228 \text{ MeV } c^{-1}$.

From $p = \gamma mv$, we have $228 \text{ MeV } c^{-1} = \gamma \times 105 \text{ MeV } c^{-2} \times v$

$$\Rightarrow \gamma \frac{v}{c} = 2.171$$

Hence

$$\frac{1}{1 - \left(\frac{v}{c}\right)^2} \left(\frac{v}{c}\right)^2 = 4.715$$

$$\Rightarrow \left(\frac{v}{c}\right)^2 = 4.715 - 4.715 \left(\frac{v}{c}\right)^2$$

$$\begin{aligned} \Rightarrow \left(\frac{v}{c}\right)^2 &= \frac{4.715}{5.715} \\ &= 0.8250 \end{aligned}$$

and so $v = 0.91c$.

Exam tip

You can also do this by first finding the total energy (251 MeV) and then the gamma factor (2.39) and then the speed.

A.28 Find the kinetic energy of an electron whose momentum is $1.5 \text{ MeV } c^{-1}$.

The total energy of the electron is given by

$$\begin{aligned} E^2 &= m^2 c^4 + p^2 c^2 \\ &= 0.511 \text{ MeV}^2 + 1.52 \text{ MeV}^2 c^{-2} \times c^2 \\ &= 2.511 \text{ MeV}^2 \end{aligned}$$

$$\Rightarrow E = 1.58 \text{ MeV}$$

and so

$$\begin{aligned} E_k &= E - mc^2 \\ &= 1.58 \text{ MeV} - 0.511 \text{ MeV} \\ &= 1.07 \text{ MeV} \end{aligned}$$

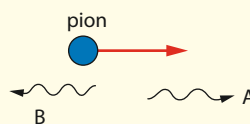
Nature of science

General principles survive paradigm shifts – usually

Newtonian mechanics relies on basic quantities such as mass, energy and momentum and the laws that relate these quantities. Einstein realised that laws such as energy and momentum conservation are the result of deeper principles and sought to preserve them in relativity. This meant modifying what we normally call energy and momentum in such a way that these quantities are also conserved in relativity.

? Test yourself

- 42 Calculate the energy needed to accelerate an electron to a speed of
- $0.50c$
 - $0.90c$
 - $0.99c$.
- 43 **a** Calculate the speed of a particle whose kinetic energy is 10 times its rest energy.
b Calculate the speed of a particle whose total energy is 10 times its rest energy.
- 44 Find the momentum of a proton whose total energy is 5 times its rest energy.
- 45 Find the total energy of an electron with a momentum of $350 \text{ MeV } c^{-1}$.
- 46 Determine the momentum, in conventional SI units, of a proton with momentum $685 \text{ MeV } c^{-1}$.
- 47 Find the kinetic energy of a proton whose momentum is $500 \text{ MeV } c^{-1}$.
- 48 The rest energy of a particle is 135 MeV and its total energy is 200 MeV . Find its speed.
- 49 Calculate the potential difference through which a proton must be accelerated (from rest) so that its momentum is $1200 \text{ GeV } c^{-1}$.
- 50 Find the momentum of a proton whose speed is $0.99c$.
- 51 Calculate the speed of a proton with momentum $1.5 \text{ GeV } c^{-1}$.
- 52 A proton initially at rest finds itself in a region of uniform electric field of magnitude $5.0 \times 10^6 \text{ V m}^{-1}$. The electric field accelerates the proton over a distance of 1.0 km . Calculate:
- the kinetic energy of the proton
 - the speed of the proton.
- 53 **a** Show that the speed of a particle with rest mass m and momentum p is given by
- $$v = \frac{pc^2}{\sqrt{m^2c^4 + p^2c^2}}$$
- b** An electron and a proton have the same momentum. Calculate the ratio of their speeds when the momentum is
- $1.00 \text{ MeV } c^{-1}$
 - $1.00 \text{ GeV } c^{-1}$.
- c** What happens to the value of the ratio as the momentum gets larger and larger?
- 54 A particle at rest breaks apart into two pieces with masses $250 \text{ MeV } c^{-2}$ and $125 \text{ MeV } c^{-2}$. The lighter fragment moves away at a speed of $0.85c$.
- Find the speed of the other fragment.
 - Determine the rest mass of the original particle that broke apart.
- 55 An electron and a positron, each with kinetic energy 2.0 MeV , collide head-on. The electron and positron annihilate each other, giving two photons.
- Explain why the electron–positron pair cannot create just one photon.
 - Explain why the photons must be moving in opposite directions.
 - Calculate the energy of each photon.
- 56 A neutral pion of mass $135 \text{ MeV } c^{-2}$ travelling at $0.80c$ decays into two photons travelling in opposite directions, $\pi^0 \rightarrow 2\gamma$. Calculate the ratio of the frequency of photon A to that of photon B.



- 57 Two identical bodies with rest mass 3.0 kg are moving directly towards one another, each with a speed of $0.80c$ relative to the laboratory. They collide and form one body. Determine the rest mass of this body.
- 58 State the formulas, in terms of the rest mass m of a particle, for
- the relativistic momentum p
 - the total energy E .
- c** Using these formulas, derive the formula $v = \frac{pc^2}{E}$ for the speed v of the particle.
- d** The formula in **c** applies to all particles, even those that are massless. Deduce that, if the particle is a photon, then $v = c$.
- 59 **a** Show that when a particle of mass m and charge q is accelerated from rest through a potential difference V , the speed it attains corresponds to a gamma factor $\gamma = 1 + \frac{qV}{mc^2}$.
- b** A proton is accelerated from rest through a potential difference V . Calculate V such that the proton reaches a speed of $0.998c$.
- 60 Show that a free electron cannot absorb or emit a photon.



A5 General relativity (HL)

The theory of general relativity was formulated by Albert Einstein in 1915. It is a theory of gravitation that replaces the standard theory of gravitation of Newton, and generalises Einstein's special theory of relativity. The theory of general relativity stands as the crowning achievement of Einstein's genius and is considered to be perhaps the most elegant and beautiful example of a physical theory ever constructed. It is a radical theory in that it relates the distribution of matter and energy in the universe to the structure of space and time. The geometry of spacetime is a direct function of the matter and energy that spacetime contains.

A5.1 The principle of equivalence

We saw in Topic 2 that when a person stands on a scale inside a freely falling elevator ('Einstein's elevator') the reading of the scale is zero. It is as if the person is weightless. This is what the scale would read if the elevator were moving at constant velocity in deep space, far from all masses. Similarly, consider an astronaut in a spacecraft in orbit around the Earth. She too feels weightless and floats inside the spacecraft. But neither the person in the falling elevator nor the astronaut is really weightless. Gravity does act on both. We can say that the acceleration of the freely falling elevator or the centripetal acceleration of the spacecraft has 'cancelled out' the force of gravity. The right acceleration can make gravity 'disappear' and make the frame of reference under consideration look like one moving at constant velocity.

The right acceleration can also make gravity 'appear'. Consider an astronaut in a spacecraft that is moving with constant velocity in deep space, far from all masses. The astronaut really is weightless. The spacecraft engines are now ignited and the spacecraft accelerates forward. The astronaut feels pinned down to the floor. If he drops a coin, it will hit the floor, whereas previously it would have floated in the spacecraft. The coin falling to the floor and the sensation of being pinned down are what we normally associate with gravity.

These are two examples where the effects of acceleration mimic those of gravity. Another expression for 'the effects of acceleration' is 'inertial effects'. Einstein elevated these observations to a principle of physics – the **equivalence principle** (EP):

Gravitational and inertial effects are indistinguishable.

Applying this principle to the two examples we discussed above, we may re-express it more precisely in two versions:

Version 1: A reference frame moving at constant velocity far from all masses is equivalent to a freely falling reference frame in a uniform gravitational field.

Version 2: An accelerating reference frame far from all masses is equivalent to a reference frame at rest in a gravitational field.

Learning objectives

- Understand the principle of equivalence.
- Understand the bending of light.
- Describe gravitational red-shift and the Pound–Rebka experiment.
- Understand the nature of black holes.
- Understand the meaning of the event horizon.
- Understand time dilation near a black hole.
- Understand the implications of general relativity for the universe as a whole.

Exam tip

There is hardly an examination paper in which you will not be asked to state this principle.

Exam tip

You can use either of these versions of the equivalence principle, but the original statement takes care of both.

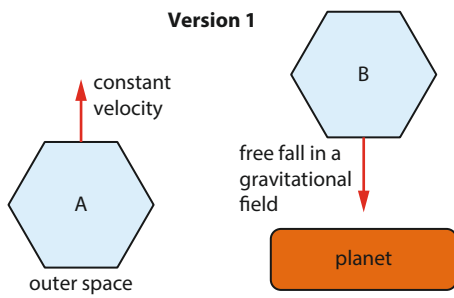


Figure A.44 A frame of reference moving at constant velocity far from any masses (A) and a freely falling frame of reference in a gravitational field (B) are equivalent.

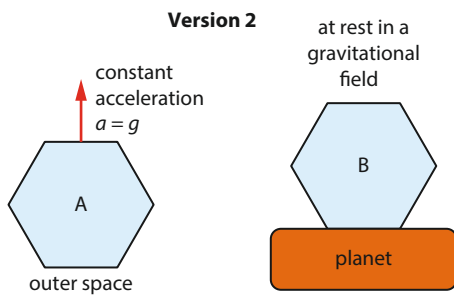


Figure A.45 An accelerating frame of reference far from any masses (A) and a frame at rest in a gravitational field (B) are equivalent.

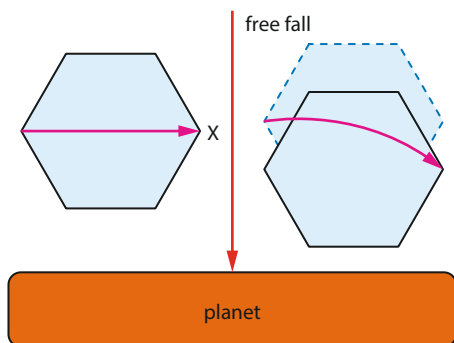


Figure A.46 A ray of light bends towards a massive object.

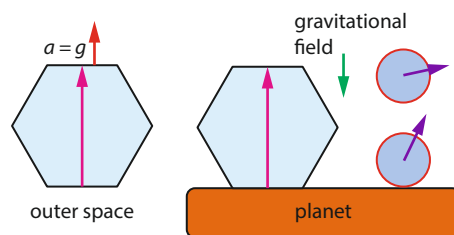


Figure A.47 In the accelerated frame the light ray will have a lower frequency at the top because of the Doppler effect.

Exam tip

In an exam you should be able to use the EP to deduce the bending of light, gravitational red-shift and time dilation.

Figure A.44 shows a frame of reference A moving at constant velocity in deep space, and a frame of reference B falling freely in a gravitational field. The EP says that the two frames are equivalent. There is no experiment that the occupants of A can perform that will give a different result from the same experiment performed in B, nor can the occupants of either frame decide which of the two states of motion they are ‘really’ in.

Figure A.45 shows a frame of reference A accelerating in deep space, and a frame of reference B at rest in a gravitational field. Again, there is no experiment that the occupants of A can perform that will give a different result from the same experiment performed in B, nor can the occupants decide which is which.

This principle has immediate consequences.

A5.2 Consequences of the EP: the bending of light

One consequence of the EP is that **light bends towards a massive body**. Imagine a box falling freely in the gravitational field of a planet (Figure A.46). A ray of light is emitted from inside the box initially parallel to the floor. Since, by the EP, this frame is equivalent to a frame moving at constant velocity in outer space, there can be no doubt that the ray of light will hit the opposite side of the box at point X. But an observer, (who must also see the ray hit at X), says that the ray is bending towards the planet.

But this means that, from the point of view of an observer at rest on the surface of the massive body, light has bent towards it.

A ray of light bends towards a massive body.

Massive objects can thus act as a kind of **gravitational lens**.

A5.3 Consequences of the EP: gravitational red-shift

A second consequence of the EP is that a ray of light emitted upwards in a gravitational field will have its frequency reduced as it climbs higher. In Figure A.47 a reference frame is at rest in a gravitational field of a planet. A ray of light is emitted upwards. This frame is equivalent to a frame in outer space with an acceleration equal to the gravitational field strength on the planet. In this accelerated frame, an observer at the top of the frame is moving away from the light emitted at the base. So, according to the Doppler effect, he will observe a lower frequency than that emitted at the base. By the EP, the same thing happens in a frame at rest in a gravitational field.

The frequency of a ray of light is reduced as it moves higher in a gravitational field.

Notice that if the frequency decreases as the ray moves upwards, the period must increase (the speed of light is constant). But the period may be used as a clock. The period is the time in between ticks of a clock. **Gravitational red-shift** is therefore equivalent to saying that we have a gravitational time dilation effect. Consider two identical clocks.



One is placed near a massive body and the other far from it. When the clock near the massive body shows that 1 s has gone by, the faraway clock will show that more than 1 s has gone by.

Time slows down near a massive body.

A5.4 The tests of general relativity

There are many experimental tests that the general theory of relativity has passed. These include:

- The **bending of light** and radio signals near massive objects: this was measured by Eddington in 1919, in agreement with Einstein's prediction.
- Gravitational frequency shift: this was observed by Pound and Rebka in 1960 (see Section A5.5).
- The Shapiro time delay experiment: radio signals sent to a planet or spacecraft and reflected back to the Earth take slightly longer to return when they pass near the Sun than when the Sun is out of the way. This delay is predicted by the theory of relativity.
- The precession of the perihelion of Mercury: Mercury has an abnormality in its orbit that Newtonian gravitation cannot account for. Einstein's theory correctly predicts this anomaly.
- Gravitational lensing: massive galaxies may bend light from distant stars and even galaxies, creating multiple images. This has been observed; see Figure A.48.



Figure A.48 Multiple images of a distant quasar, caused by an intervening galaxy whose gravitational field has deflected the light from the quasar.

A5.5 The Pound–Rebka experiment

The phenomenon of gravitational red-shift was experimentally verified by the **Pound–Rebka experiment** in 1960. In this experiment, performed at Harvard University, a beam of gamma rays of energy 14.4 keV from a nuclear transition in iron-57 was emitted from the top of a tower 22.6 m high and detected at ground level; see Figure A.49. Just as the frequency of light decreases as the light moves upwards in a gravitational field, it increases if the light is moving downwards: we have a blue-shift (which is what Pound and Rebka measured).

Theory predicts that the frequency shift Δf between the emitted and received frequencies is given by

$$\frac{\Delta f}{f} = \frac{gH}{c^2}$$

Here, f stands for the emitted or the observed frequency and H is the height from which the photons are emitted.

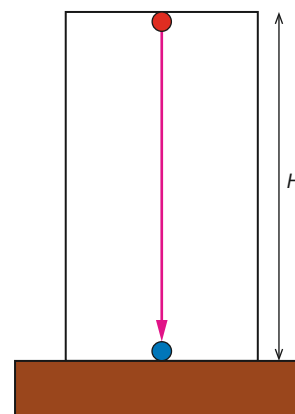


Figure A.49 A photon is emitted from the top of a tower and observed at the base, where its frequency is measured and found to be higher than that emitted.

Worked example

A.29 A photon of energy 14.4 keV is emitted from the top of a 30 m-tall tower towards the ground. Calculate the shift in frequency expected at the base of the tower.

On emission, the photon has a frequency given by

$$\begin{aligned}E &= hf \\ \Rightarrow f &= \frac{E}{h} \\ &= \frac{14.4 \times 10^3 \times 1.6 \times 10^{-19}}{6.63 \times 10^{-34}} \\ &= 3.475 \times 10^{18} \text{ Hz}\end{aligned}$$

The shift (it is a blue-shift) is thus

$$\begin{aligned}\frac{\Delta f}{f} &= \frac{gH}{c^2} \\ \Rightarrow \Delta f &= f \frac{gH}{c^2} \\ &= \frac{10 \times 30}{9 \times 10^{16}} \times 3.475 \times 10^{18} \text{ Hz} \\ &= 1.16 \times 10^4 \text{ Hz}\end{aligned}$$

Note how sensitive this experiment actually is: the shift in frequency is only about 10^4 Hz, compared with the emitted frequency of about 10^{18} Hz, a fractional shift of only

$$\frac{10^4 \text{ Hz}}{10^{18} \text{ Hz}} = 10^{-14}$$



Credit where credit is due

Einstein's general theory of relativity would have been impossible to develop were it not for 19th-century mathematicians who were bold enough to question Euclid's fifth axiom of geometry. Modifying that axiom meant that new kinds of geometry became available. Developing the mathematical machinery to describe these geometries was one of the greatest achievements of 19th-century mathematics.

A5.6 The structure of the theory

The theory of general relativity is a physical theory different from all others, in that:

The mass and energy content of space determines the geometry of that space and time. The geometry of spacetime determines the motion of mass and energy in the spacetime.

That is,

$$\text{geometry} \Leftrightarrow \text{mass-energy}$$

Spacetime is a four-dimensional world with three space coordinates and one time coordinate. In the absence of any forces, a body moves in this four-dimensional world along paths of shortest length, called **geodesics**.

If a single mass M is the only mass present in the universe, the solutions of the Einstein equations imply that, far from this mass, the geometry of space is the usual Euclidean flat geometry, with all its familiar rules (for example, the angles of a triangle add up to 180°). As we approach the neighbourhood of M , the space becomes curved, as illustrated in

Figure A.50. The rules of geometry then have to change. Large masses with small radii produce extreme bending of the spacetime around them. Thus, the motion of a planet around the Sun is, according to Einstein, not the result of a gravitational force acting on the planet (as Newton would have it) but rather due to the curved geometry in the space and time around the Sun created by the large mass of the Sun.

Similarly, light retains its familiar property of travelling from one place to another in the shortest possible time. Since the speed of light is constant, this means that light travels along paths of shortest length: geodesics. In the flat Euclidean geometry we are used to, geodesics are straight lines. In the curved non-Euclidean geometry of general relativity, something else replaces the concept of a straight line. A ray of light travelling near the Sun looks bent to us because we are used to flat space. But the geometry near the Sun is curved and the ‘bent’ ray is actually a geodesic: it is the ‘straight’ line appropriate to that geometry.

A5.7 Black holes

The theory of general relativity also predicts the existence of objects that contract under the influence of their own gravitation, becoming ever smaller. No mechanism is known for stopping this collapse, and the object is expected to become a hole in spacetime, a point of infinite density. This creates an immense bending of spacetime around this point. This point is called a **black hole**, a name coined by John Archibald Wheeler (Figure A.51), since nothing can escape from it. Massive stars can, under appropriate conditions, collapse under their own gravitation and end up as black holes (see Option D, Astrophysics). Powerful theorems by Stephen Hawking and Roger Penrose show that the formation of black holes is inevitable and not dependent too much on the details of how the collapse itself proceeds.

A5.8 The Schwarzschild radius and the event horizon

A distance known as the **Schwarzschild radius** of a black hole is of importance in understanding the behaviour of black holes. Karl Schwarzschild (Figure A.52) was a German astronomer who provided the first solution of the Einstein equations.

The Schwarzschild radius is given by

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

where M is the mass and c the speed of light. This radius is not the actual radius of the black hole (the black hole is a point), but the distance from the hole’s centre that separates space into a region from which an object can escape and a region from which no object can escape (Figure A.53).

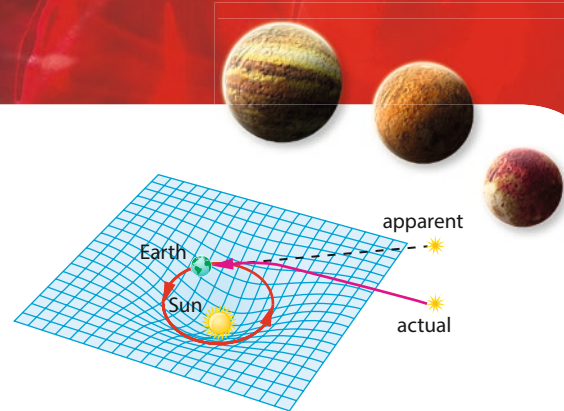


Figure A.50 The mass of the Sun causes a curvature of the space around it. Light from a star bends on its way to the Earth and this makes the star appear to be in a different position from its actual position.

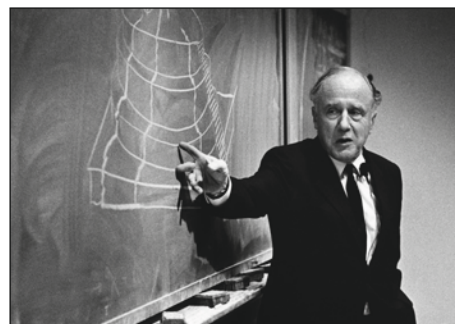


Figure A.51 John A. Wheeler, a legendary figure in black-hole physics and the man who coined the term.

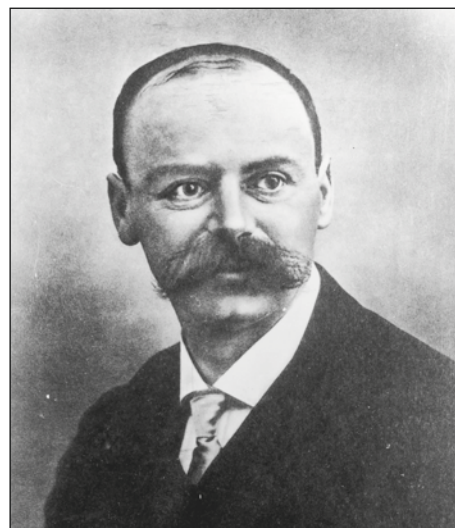


Figure A.52 Karl Schwarzschild was the man who provided the first solution of Einstein’s equations.

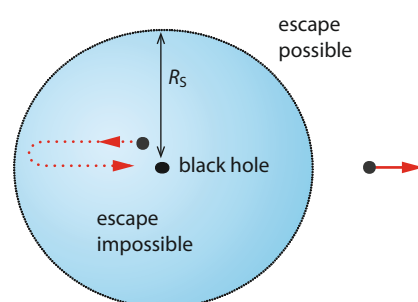


Figure A.53 The Schwarzschild radius splits space into two regions. From within this radius nothing – not even light – can escape.



Quantum black holes are different from classical black holes

The statement that nothing escapes from a black hole is true only if we ignore quantum effects. Stephen Hawking showed in 1974 that, when quantum effects are taken into account, black holes radiate just like a black body does, with the reciprocal of the mass playing the role of temperature. For massive black holes this is not significant, since in that case the temperature is very small. However, it is interesting that, once the concept of ‘temperature’ was introduced, knowledge of thermodynamics could be transferred to black-hole physics in a formal way, even though black holes do not have a temperature in the normal sense.

Any object closer to the centre of the black hole than R_S will fall into the hole; no amount of energy supplied to this body will allow it to escape from the black hole.

The escape velocity at a distance R_S from the centre of the black hole is the speed of light, so nothing can escape from within this radius. The Schwarzschild radius is also called the **event horizon radius**. The latter name is apt, since anything taking place within the event horizon cannot be seen by or communicated to the outside. The area of the event horizon is taken as the surface area of the black hole (but, remember, the black hole is a point).

The Schwarzschild radius can be derived in an elementary way without recourse to general relativity if we assume that a photon has a mass m on which the black hole’s gravity acts. Then, as in the calculation of escape velocity in Topic 10 on gravitation,

$$\frac{1}{2}mc^2 = \frac{GM}{R_S}m$$

Thus m cancels and we can solve for R_S .

It can be readily calculated from this formula that, for a star of one solar mass ($M \approx 2 \times 10^{30}$ kg), the Schwarzschild radius is about 3 km:

$$\begin{aligned} R_S &= \frac{2GM}{c^2} \\ &= \frac{2 \times 6.67 \times 10^{-11} \times 2 \times 10^{30}}{(3 \times 10^8)^2} \\ &\approx 3 \times 10^3 \text{ m} \end{aligned}$$

This means, for example, that if the Sun were to become a black hole its entire mass would be confined within a sphere with a radius of 3 km or less. For a black hole of one Earth mass, this radius would be just 9 mm.

Figure A.54 shows that an observer on the surface of a star that is about to become a black hole would only be able to receive light through a cone that gets smaller and smaller as the star approaches its Schwarzschild radius. This is because rays of light coming ‘sideways’ will bend towards the star and will not reach the observer.

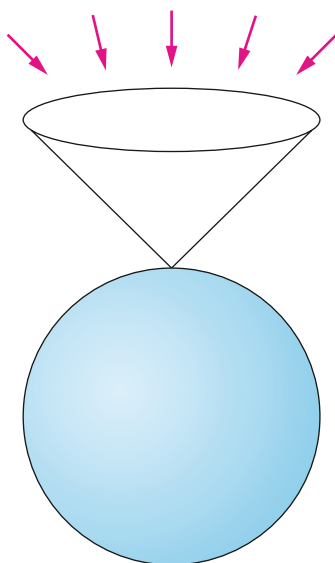


Figure A.54 The event horizon on a collapsing star is rising; that is, only rays within a steadily shrinking vertical cone can reach an observer on the star.

A5.9 Time dilation in general relativity

We have seen that the EP predicts that time runs slow near massive bodies. A special case of this phenomenon takes place near a black hole.

Consider a clock that is at a distance r from the centre of a black hole of Schwarzschild radius R_S (the clock is outside the event horizon, i.e. $r > R_S$); see Figure A.55. An observer who is stationary with respect to the clock measures the time interval between two ticks of the clock to be Δt_0 . An identical clock very far away from the black hole will measure a time interval

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - \frac{R_S}{r}}}$$

This formula is the general relativistic analogue of time dilation in special relativity. It applies near a black hole.

Thus, consider a (theoretical) observer approaching a black hole. This observer sends signals to a faraway observer in a spacecraft, informing the spacecraft of his position. When his distance from the centre of the black hole is $r = 1.50R_S$, the observer stops and sends two signals 1 s apart (as measured by his clocks). The spacecraft observers will receive signals separated in time by

$$\begin{aligned} \Delta t &= \frac{\Delta t_0}{\sqrt{1 - \frac{R_S}{r}}} \\ &= \frac{1.00}{\sqrt{1 - \frac{1}{1.50}}} \\ &= 1.73 \text{ s} \end{aligned}$$

The extreme case of time dilation is when the observer is just an infinitesimally small distance outside the event horizon. If he stops there and sends two signals 1 s apart, the faraway observer will receive the signals separated by an enormous interval of time. In particular, if the observer is *at* the event horizon when the second signal is emitted, that signal will never be received.

A5.10 General relativity and the universe

Soon after he published his general theory of relativity, Einstein applied his equations for general relativity to the universe as a whole. The result looks like this:

$$R_{\mu\nu} - \left(\frac{1}{2}R - \Lambda\right)g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}$$

and will not be examined!

On a large scale, the universe looks like a cloud of dust of density ρ . Under various simplifying assumptions, the equations reduce to an equation for a quantity that we will loosely call the ‘radius’ of the universe, R . Solving these equations gives the dependence of R on time. Einstein himself believed in a static universe, which would have $R = \text{constant}$. His calculations, however, did not give a constant R .

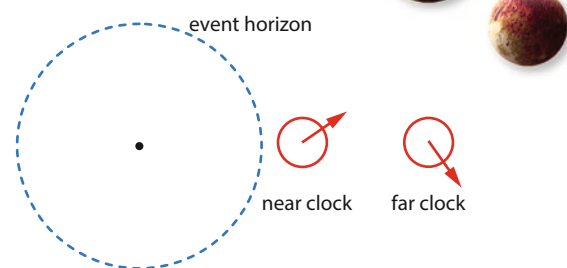


Figure A.55 A clock near a black hole runs slowly compared with a clock far away.

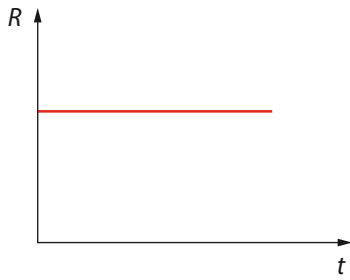


Figure A.56 A model of the universe with a constant radius. Einstein introduced the cosmological constant in order to make the universe static.

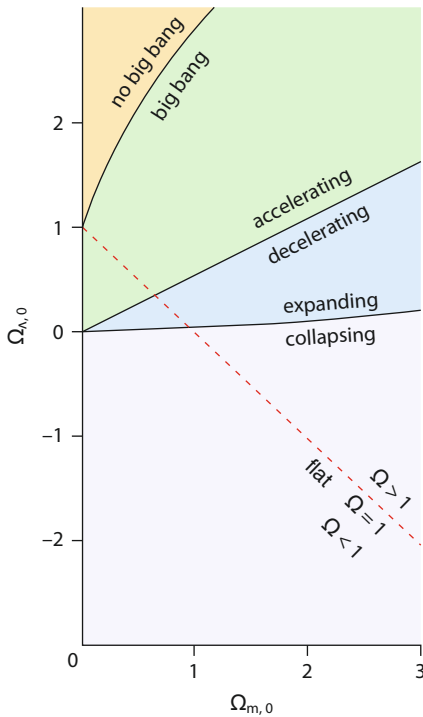


Figure A.57 There are various possibilities for the evolution of the universe, depending on how much energy and mass it contains. (From M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004. © The Open University, used with permission)

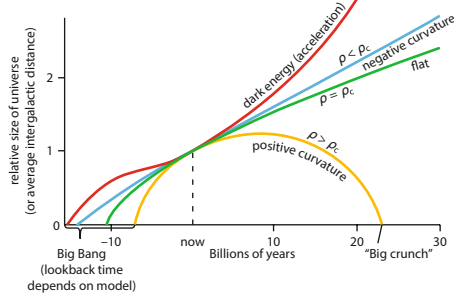


Figure A.59 Solutions of Einstein's equations for the evolution of the scale factor. The present time is indicated by 'now'. Notice that the age of the universe varies depending on which solution is chosen – in other words, different solutions imply different ages. Three models assume zero dark energy; the red line does not.

So he modified them, adding the famous **cosmological constant** term Λ . This term made R constant! This is shown in Figure A.56.

In this model there is no 'Big Bang' and the universe always has the same size. This was before Hubble discovered that the universe is expanding. Einstein missed the chance to theoretically predict an expanding universe; he later called the addition of the cosmological constant 'the greatest blunder of my life'. This constant may be thought of as being related to a vacuum energy, energy that is present in all space. This energy is now called **dark energy**. The cosmological constant went into obscurity for many decades but it did not go away: it was to make a comeback with a vengeance much later!

The first serious attempt to find how R depends on time was made by the Russian mathematician Alexander Friedmann (1888–1925), in work later taken up by Lemaître, Robertson and Walker. Friedmann applied the Einstein equations and realised that there were a number of possibilities: the solutions depend on how much matter and energy the universe contains (Figure A.73).

We define the **density** parameter for matter, Ω_m , and for dark energy, Ω_Λ , as

$$\Omega_m = \frac{\rho_m}{\rho_{\text{crit}}} \quad \text{and} \quad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{\text{crit}}}$$

where ρ_m is the actual density of matter in the universe and ρ_{crit} is a reference density called the critical density, about $10^{-26} \text{ kg m}^{-3}$; ρ_Λ is the density of dark energy. The Friedmann equations give various solutions, depending on the values of Ω_m and Ω_Λ . Deciding which solution to pick depends crucially on these values, which is why cosmologists have expended enormous amounts of energy and time trying to accurately measure Ω_m and Ω_Λ . In Figure A.57 the subscript 0 indicates the values of these parameters at the present time. There are four main regions in this diagram. Models above the red dashed line are ones in which the geometry of the universe resembles the surface of a sphere (Figure A.58c). Points below the line have a geometry like that of the surface of a saddle (Figure A.58a). Points on the line imply a flat universe where the rules of Euclidean geometry apply (Figure A.58b).

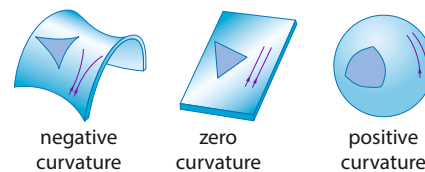


Figure A.58 Three models of the universe with different curvature. **a** In negative-curvature models, the angles of a triangle add up to less than 180° and initially parallel lines eventually diverge. **b** Ordinary flat geometry. **c** Positive curvature, in which the angles of a triangle add up to more than 180° and initially parallel lines eventually intersect. (From M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004. © The Open University, used with permission)

Figure A.59 shows how the radius R of the universe varies with time for various values of the parameters.

In all cases the radius starts from zero, implying a 'Big Bang'. In one possibility, $R(t)$ starts from zero, increases to a maximum value and then decreases to zero again – the universe collapses after an initial period of expansion. This is called the **closed** model, and corresponds to $\Omega_m > 1$,



that is, $\rho_m > \rho_{\text{crit}}$. A second possibility applies when $\Omega_m < 1$, that is, $\rho_m < \rho_{\text{crit}}$. The present data from the Planck satellite observatory indicate that $\Omega_m \approx 0.32$ and $\Omega_\Lambda \approx 0.68$, so $\Omega_m + \Omega_\Lambda \approx 1$, on the red dashed line in Figure A.57. This implies a flat universe, meaning that at present our universe has a flat geometry, 32% of its mass–energy content is matter, 68% is dark energy and it is expanding forever at an accelerating rate. This is shown by the red line in Figure A.59.

Nature of science

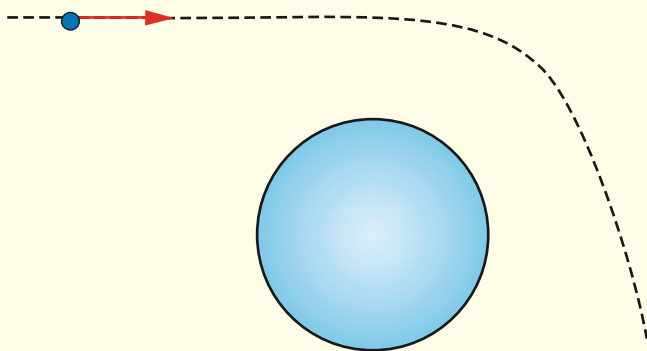
Creative and critical thinking

The theory of general relativity is an almost magical theory in which the energy and mass content of spacetime determines the geometry of spacetime. In turn, the kind of geometry spacetime has determines how the mass and energy of the spacetime move about. To connect these ideas in the theory of general relativity, Einstein used intuition, creative thinking and imagination. Initial solutions of Einstein's equations suggested that no light could escape from a black hole. Then a further imaginative leap – the development of quantum theory – showed an unexpected connection between black holes and thermodynamics.

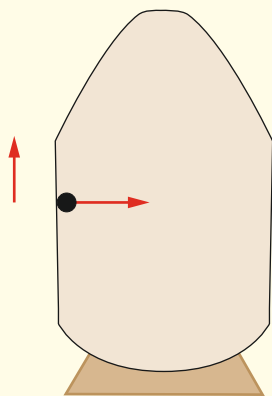
? Test yourself

- 61 Discuss the statement 'a ray of light does not actually bend near a massive object but follows a straight-line path in the geometry of the space around the massive object'.
- 62 A spacecraft filled with air at ordinary density and pressure is far from any large masses. A helium-filled balloon floats inside the spacecraft, which now accelerates to the right. Determine which way (if any) the balloon moves.
- 63 The spacecraft of question 62 now has a lighted candle in it. The craft accelerates to the right. Discuss what happens to the flame of the candle.
- 64
 - a Describe what is meant by the equivalence principle.
 - b Explain how this principle leads to the predictions that:
 - i light bends in a gravitational field
 - ii time 'runs slower' near a massive object.
- 65 Describe what the general theory of relativity predicts about a massive object whose radius is getting smaller.
- 66 In a reference frame falling freely in a gravitational field, an observer attaches a mass m to the end of a spring, extends the spring and lets the mass go. He measures the period of oscillation of the mass. Discuss whether he will find the same answer as an observer doing the same thing:
 - a on the surface of a very massive star
 - b in a true inertial frame far from any masses.
- 67 Calculate the shift in frequency of light of wavelength 500.0 nm emitted from sea level and detected at a height of 50.0 m.
- 68 A collapsed star has a radius that is 5.0 times larger than its Schwarzschild radius. An observer on the surface of the star carries a clock and a laser. Every second, the observer sends a short pulse of laser light of duration 1.00 ms and wavelength 4.00×10^{-7} m (as measured by her instruments) towards another observer in a spacecraft far from the star. Discuss qualitatively what the observer in the spacecraft can expect to measure for:
 - a the wavelength of the pulses
 - b the frequency of reception of consecutive pulses
 - c the duration of the pulses.

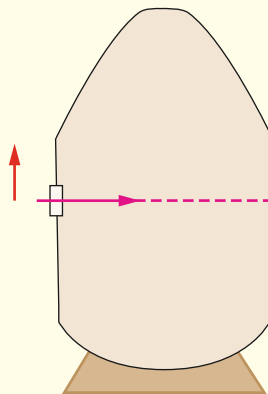
- 69 Two identical clocks are placed on a rotating disc. One is placed on the circumference of the disc and the other halfway towards the centre. Explain why the clock on the circumference will run slowly relative to the other clock.
- 70 Calculate the density of the Earth if its entire mass were confined within a radius equal to its Schwarzschild radius.
- 71 Calculate the Schwarzschild radius of a black hole with a mass equal to 10 solar masses.
- 72 Explain what is meant by **geodesic**.
- 73 A mass m moves past a massive body along the path shown below. Explain the shape of the path according to:
- Newtonian gravity
 - Einstein's theory of general relativity.



- 74 A ball is thrown with a velocity that is initially parallel to the floor of a spacecraft, as shown below. Draw and explain the shape of the ball's path when:
- the spacecraft is moving with constant velocity in deep space, far from any large masses
 - the spacecraft is moving with constant acceleration in deep space, far from any large masses.



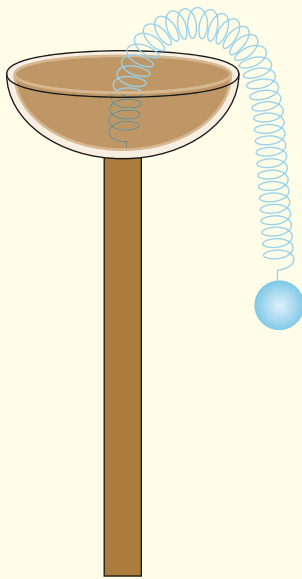
- 75 A ray of light parallel to the floor of a spacecraft enters the spacecraft through a small window, as shown below. Draw and explain the shape of the ray's path when:
- the spacecraft is moving with constant velocity in deep space, far from any large masses
 - the spacecraft is moving with constant acceleration in deep space, far from any large masses.



- 76 An observer is standing on the surface of a massive object that is collapsing and is about to form a black hole. Describe what the observer sees in the sky:
- before the object shrinks past its event horizon
 - after the object goes past its event horizon.
- 77 A plane flying from southern Europe to New York City will fly over Ireland, across the North Atlantic, over the east coast of Canada and then south to New York. Suggest why such a path is followed.

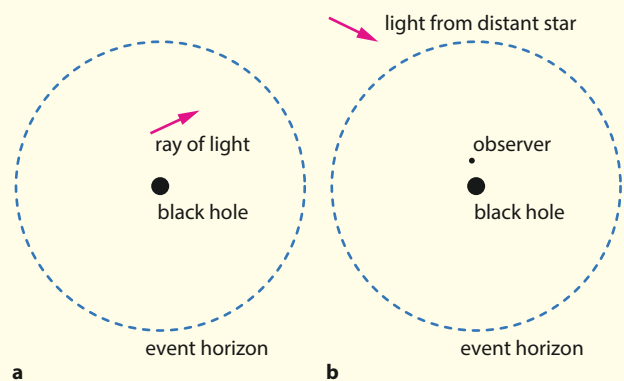


78 Einstein's birthday present. A colleague of Einstein at Princeton presented him with the following birthday present. A long tube was connected to a bowl at its top end. A spring was attached to the base of the bowl and connected to a heavy brass ball that hung out of the bowl. The spring was very 'weak' and could not pull the ball into the bowl. The exercise was to find a sure-fire method for putting the ball into the bowl without touching it. Einstein immediately found a way of doing it. Can you?



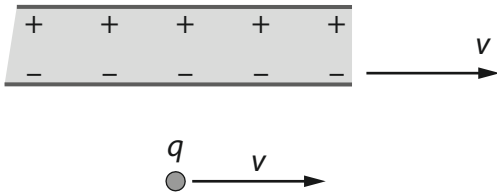
- 79** An observer approaching a black hole stops and sends signals to a faraway spacecraft every 1.0s as measured by his clocks. The signals are received 2.0s apart by observers in the spacecraft. Determine how close he is to the event horizon.
- 80** A spacecraft accelerates in the vacuum of outer space far from any masses. An observer in the spacecraft sends a radio message to a stationary spacecraft far away. The duration of the radio transmission is 5.0s, according to the observer's clock in the moving spacecraft. Explain whether the transmission will take less than, exactly or more than 5.0s when it is received by the faraway stationary spacecraft.
- 81** A black hole has a mass of 5.00×10^{35} kg.
- State what is meant by a **black hole**.
 - Calculate the Schwarzschild radius R_S of the black hole.
 - Explain why the Schwarzschild radius is not in fact the actual radius of the black hole.

- Blue light of frequency 7.50×10^{14} Hz is emitted by a source that is stationary at a distance of $0.10 R_S$ above the event horizon of a black hole. Calculate the period of this blue light according to an observer next to the source.
 - Determine the frequency measured by a distant observer who receives light emitted by this source.
- 82**
- State the formula for the gravitational (Schwarzschild) radius R_S of a black hole of mass M .
 - The sphere of radius R_S around a black hole is called the **event horizon**. State the area of the event horizon of a black hole of mass M .
 - Suggest why, over time, the area of the event horizon of a black hole always increases.
 - In physics, there is one other physical quantity that always increases with time. Can you state what that quantity is? (You will learn a lot of interesting things if you pursue the analogy implied by your answers to **c** and **d**.)
- 83**
- A ray of light is emitted from within the event horizon (dashed circle) of a black hole, as shown below. Copy the diagram and draw a possible path for this ray of light.
 - Light from a distant star arrives at a theoretical observer within the event horizon of a black hole, as shown in part **b** of the figure. Explain how it is possible for the ray shown to enter the observer's eye.



Exam-style questions

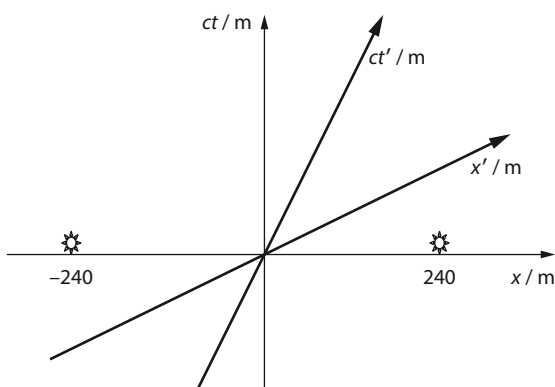
- 1 a State what is meant by a **reference frame**. [2]
- b Discuss the speed of light in the context of Maxwell's theory. [2]
- c A positive charge q moves parallel to a wire that carries electric current. The velocity of the charged particle is the same as the drift velocity of the electrons in the wire.



Discuss the force experienced by the charged particle from the point of view of an observer:

- i at rest with respect to the wire. [2]
- ii at rest with respect to the charged particle. [2]
- 2 A rocket moves past a space station at a speed of $0.98c$. The proper length of the space station is 480 m. The origins of the space station and rocket frames coincide when the rocket is moving past the middle of the space station. At that instant, clocks in both frames are set to zero.
- a State what is meant by **proper length**. [1]
- b Determine the length of the space station according to an observer in the rocket. [2]
- c Two lamps at each end of the space station turn on simultaneously according to space station observers, at time zero.
- Determine
- i which lamp turns on first, according to an observer in the rocket [4]
- ii the time interval between the lamps turning on, according to an observer in the rocket. [3]

- d The spacetime diagram below applies to the frames of the space station and the rocket. The thin axes represent the space station reference frame.



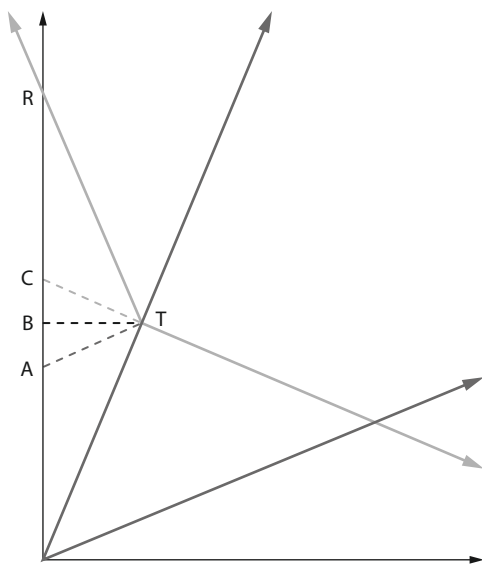
Use the diagram to verify your answer to c i. [1]

- e By drawing appropriate lines on a copy of the spacetime diagram, show the events:
- i 'light from the lamp at $x = 240$ m reaches the rocket' (the rocket is assumed to be a point) [2]
- ii 'light from the lamp at $x = -240$ m reaches the rocket' (the rocket is assumed to be a point). [2]



- 3 A rocket moves to the right with speed $0.78c$ relative to a space station. The rocket emits two missiles, each at a speed of $0.50c$ relative to the rocket. One missile, R, is emitted to the right and the other, L, to the left.
- a Calculate the velocity of missile
 - i R relative to the space station [2]
 - ii L relative to the space station [2]
 - iii R relative to missile L. [2]
 - b Show by explicit calculation that the speed of light is independent of the speed of its source. [2]

- 4 The spacetime diagram below may be used to discuss the twin paradox. The thin black axes represent the reference frame of the twin on the Earth. The dark grey axes are for the frame of the travelling twin on her way away from the Earth. The light grey axes represent her homeward frame. The time axes in each frame may be thought of as the worldlines of clocks at rest in each frame. The travelling twin moves at $0.80c$ on both legs of the trip. She turns around at event T when she reaches a planet 12ly away, as measured by the Earth observers. She returns to Earth at event R.



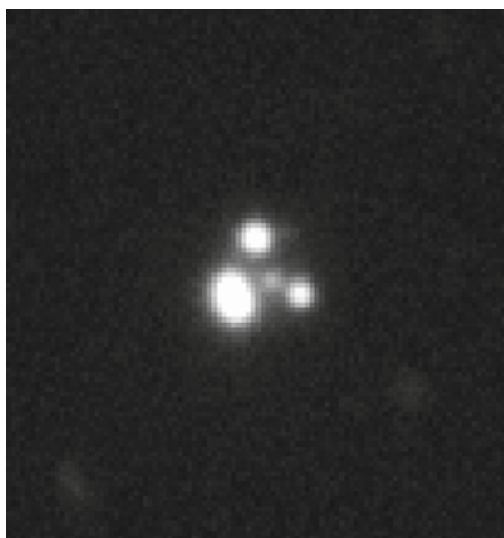
- a Describe what is meant by the **twin paradox**. [2]
- b Outline how the paradox is resolved. [2]
- c Calculate the following clock readings:
 - i the Earth clock at B [1]
 - ii the outgoing clock at T [2]
 - iii the Earth clock at A [2]
 - iv the Earth clock at C [2]
 - v the Earth clock at R [1]
 - vi the incoming clock at R. [1]
- d State by how much each twin has aged when they are reunited. [2]

- 5 A rocket of proper length 690 m moves with speed $0.85c$ past a lab. A particle is emitted from the back of the rocket and is received at the front. The particle has speed $0.75c$ as measured in the rocket frame.
- Calculate the time it takes for the particle to reach the front of the rocket according to
 - observers in the rocket [1]
 - observers in the lab. [3]
 - Suggest whether either of the times calculated in **a** is a proper time interval. [2]
 - Calculate the speed of the particle according to the lab. [2]
 - Without using Lorentz transformations, determine the distance travelled by the particle according to the lab. [2]
 - Verify your answer to **i** by using an appropriate Lorentz transformation. [2]
- 6 A spacecraft leaves the Earth at a speed of $0.60c$ towards a space station 6.0 ly away (as measured by observers on the Earth).
- Calculate the time the spacecraft will take to reach the space station, according to
 - Earth observers [1]
 - spacecraft observers. [2]
 - As the spacecraft goes past the space station it sends a radio signal back towards the Earth. Calculate the time it will take the signal to arrive on the Earth according to
 - Earth observers [2]
 - spacecraft observers. [4]
- 7 The average lifetime of a muon as measured in the muon's rest frame is about 2.2×10^{-6} s. Muons decay into electrons. A muon is created at a height of 3.0 km above the surface of the Earth (as measured by observers on the Earth) and moves towards the surface at a speed of $0.98c$. The gamma factor γ for this speed is 5.0. By appropriate calculations, explain how muon decay experiments provide evidence for
- time dilation [3]
 - length contraction. [3]
- HL** 8 **a** A proton is accelerated from rest through a potential difference of 2.5 GV. Determine, for the accelerated proton,
- the total energy [2]
 - the momentum [2]
 - the speed. [3]
- b** In a hypothetical experiment, two particles, each of rest mass $135 \text{ MeV}c^{-2}$ and moving in opposite directions with speed $0.98c$ relative to the laboratory, collide and form a single particle. Calculate the rest mass of this particle. [3]



- HL 9 a** A neutral pion with rest energy $135 \text{ MeV } c^{-2}$ and moving at $0.98c$ relative to a lab decays into two photons. One photon is emitted in the direction of motion of the pion and the other in the opposite direction.
- i** State what is meant by the **rest energy** of a particle. [1]
 - ii** Calculate the momentum and total energy of the pion as measured in the lab. [3]
- b** By applying the laws of conservation of energy and momentum to this decay, determine:
- i** the energy of the forward-moving photon [3]
 - ii** the momentum of the backward-moving photon. [2]
- c i** Show that the speed of a particle with momentum p and total energy E is given by
- $$v = \frac{pc^2}{E}. \quad [2]$$
- ii** Deduce that a particle of zero rest mass must move at the speed of light. [2]

- HL 10 a** Describe what is meant by the equivalence principle. [2]
- b i** Use the equivalence principle to deduce that a ray of light bends towards a massive body. [3]
- ii** Einstein would claim that the ray does not in fact bend. By reference to the curvature of space, suggest what Einstein might mean by this. [2]
- c** The image below shows a multiple image of a distant quasar.



The image has been formed because light from the quasar went past a massive galaxy on its way to the Earth.

Explain how the multiple image is formed. [2]

- HL 11 a** State what is meant by gravitational red-shift. [2]
- b** Explain gravitational red-shift using the equivalence principle. [4]
- c** In the Pound–Rebka experiment, a gamma ray was emitted from the top of a tower 23 m high.
- i** Calculate the fractional change in frequency observed at the bottom of the tower. [2]
 - ii** Explain why, for this experiment to be successful, frequency must be measured very precisely. [2]
 - iii** Outline why this experiment is evidence for gravitational time dilation. [2]
- d** The experiment in **c** is repeated in an elevator of height 23 m that is falling freely above the Earth. The gamma ray is emitted from the top of the elevator. Predict the frequency of the gamma ray as measured at the base of the elevator. [3]

- HL** 12 a State what is meant by:
- i** a **black hole** [1]
 - ii** the **event horizon** of a black hole. [1]
- b** Calculate the event horizon radius for a black hole of mass 5.0×10^{35} kg. [2]
- c** Suggest why the event horizon radius of a black hole is likely to increase with time. [2]
- d** A probe near the event horizon of the black hole in **b** sends signals to a spacecraft far from the hole every 5.0 s, according to clocks on the probe.
- i** The signals are received every 15 s, according to clocks on a spacecraft. Determine the distance of the probe from the black hole. [2]
 - ii** State **one other** difference that observers on the spacecraft will notice about the received signal. [1]

Option B Engineering physics

B1 Rotational dynamics

In the mechanics we have studied so far, we have assumed that we were dealing with point masses. A net force applied to a point mass will accelerate it. However, things change when we deal with extended bodies such as cylinders and spheres. Forces will not only accelerate a body's centre of mass but may also make the body rotate.

B1.1 Kinematics

Figure B.1 shows a body that rotates around an axis that is perpendicular to the plane of the paper. In time Δt the body sweeps out an angle $\Delta\theta$.

We define the average angular velocity of the body to be:

$$\bar{\omega} = \frac{\Delta\theta}{\Delta t}$$

and the instantaneous angular velocity to be:

$$\omega = \lim_{\Delta t \rightarrow 0} \frac{\Delta\theta}{\Delta t}$$

The unit of angular velocity is rad s^{-1} .

The angular velocity will increase or decrease if there is angular acceleration. We define the average angular acceleration as:

$$\bar{\alpha} = \frac{\Delta\omega}{\Delta t}$$

and the instantaneous angular acceleration as:

$$\alpha = \lim_{\Delta t \rightarrow 0} \frac{\Delta\omega}{\Delta t}$$

The unit of angular acceleration is rad s^{-2} .

We see that these definitions are completely analogous to our definitions of linear quantities: we have a 'translation' dictionary between linear and angular quantities:

Linear quantity	Angular quantity
Position, s	Angle, θ
Linear velocity, v	Angular velocity, ω
Acceleration, a	Angular acceleration, α

Precisely because of this correspondence between linear and angular quantities, where a formula applies to linear quantities a similar formula will apply to the angular quantities. Thus we have the relations

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

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

$$s = \frac{u+v}{2}t$$

$$\theta = \frac{\omega_i + \omega_f}{2}t$$

$$v = u + at$$

$$\omega_f = \omega_i + \alpha t$$

$$v^2 = u^2 + 2as$$

$$\omega_f^2 = \omega_i^2 + 2\alpha\theta$$

Learning objectives

- Understand torque.
- Apply the conditions of rotational and translational equilibrium.
- Work with the kinematic equations for rotational motion.
- Solve problems of rotational dynamics.
- Apply conservation of angular momentum.

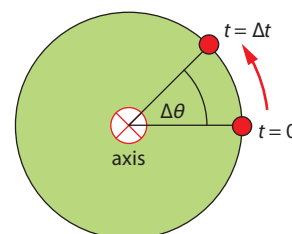


Figure B.1 A body rotating about an axis.

Exam tip

Angular velocity and angular acceleration are actually vector quantities, but we will not make use of their vector nature in this course.

Exam tip

Be careful to understand how angular acceleration α is related to linear acceleration a . This linear acceleration is the rate of change of speed, not the centripetal acceleration of circular motion.

where the subscripts i and f denote initial and final values, respectively. We saw in Topic 6 that angular velocity ω and linear velocity v are related by $v = \omega r$ (Figure B.2). A similar relation holds between angular and linear acceleration: $a = \alpha r$.

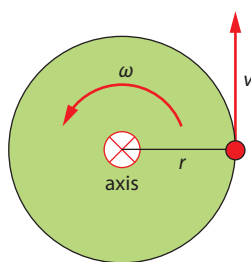


Figure B.2 Linear and angular velocities.

Similarly, whatever applies to graphs involving linear quantities also applies to graphs of angular quantities. Thus:

- in a graph of angle θ versus time t , the gradient (slope) is the angular velocity ω
- in a graph of angular velocity ω versus time t , the gradient (slope) is the **angular acceleration** and the area is the angle turned, $\Delta\theta$
- in a graph of angular acceleration α versus time t , the area is the change in angular velocity, $\Delta\omega$.

Worked examples

B.1 The initial angular speed of a rotating disc is 24 rad s^{-1} . The disc suffers an angular deceleration of 3.0 rad s^{-2} .

- Calculate how many full revolutions the disc will make before stopping.
- Determine when the disc will stop rotating.

a From $\omega^2 = \omega_i^2 + 2a\theta$ we find

$$0 = 24^2 + 2 \times (-3.0) \times \theta \Rightarrow \theta = \frac{24^2}{2 \times 3.0} = 96 \text{ rad}$$

This corresponds to $\frac{96}{2\pi} = 15.3 \approx 15$ revolutions.

b From $\omega = \omega_i + at$ we find

$$0 = 24 + (-3.0)t$$

$$t = 8.0 \text{ s}$$

B.2 A wheel of radius 0.80 m is rotating about its axis with angular speed 2.5 rad s^{-1} . Find the linear speed of a point on the circumference of the wheel.

The answer is just $v = \omega R = 2.5 \times 0.80 = 2.0 \text{ m s}^{-1}$.



B.3 Figure B.3 shows a disc of radius R rotating about its axis with constant angular velocity ω .
A point X is at the circumference of the disc and another point Y is at a distance $\frac{R}{2}$ from the axis.

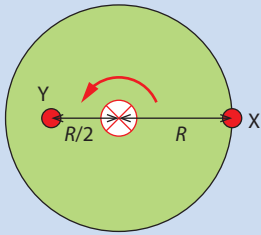


Figure B.3

Calculate the ratios

a $\frac{\omega_X}{\omega_Y}$ **b** $\frac{v_X}{v_Y}$

- a** All points on the disc have the same angular velocity, so $\frac{\omega_X}{\omega_Y} = 1$.
b From $v = \omega r$ we have that $\frac{v_X}{v_Y} = \frac{\omega R}{\omega R/2} = 2$.

B.4 The graph in Figure B.4 shows the variation of the angular velocity of a rotating body.
Calculate **a** the angular acceleration and **b** the total angle the body turns in 4.0s.



Figure B.4

- a** The angular acceleration is the gradient of the curve, so $\alpha = -3.0 \text{ rad s}^{-2}$.
b The angle turned is the area under the curve, 24rad.

B1.2 Torque

Torque refers to the ability of a force to produce rotation.

Figure B.5 shows a rigid body that is free to rotate about an axis through point P. A force F acts at point Q. The force is applied a distance r from the axis. The force makes an angle θ with the vector from P to Q. We define the torque Γ produced by the force F about the axis as the product of the force, the magnitude of r and the sine of θ :

$$\Gamma = Fr \sin \theta$$

The unit of torque is N m. Although this combination of units is equivalent to J (joules), by convention torque is never expressed in J.

Exam tip

The angle θ (is the angle between the vector r and the force F).

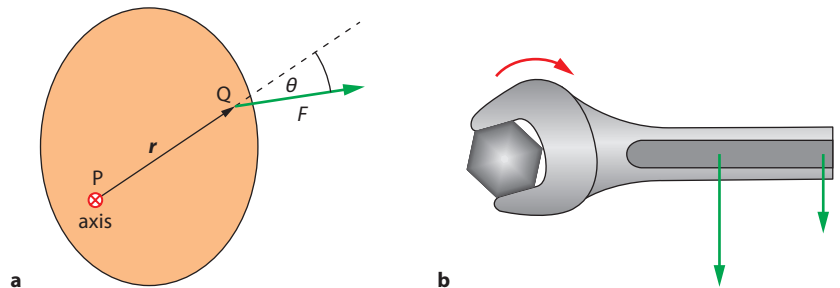


Figure B.5 a The force tends to rotate the body in a clockwise direction about an axis out of the page. This is because the force produces a torque about that axis.

b A spanner turning a screw: a real-life application of torque. A smaller force further from the axis would have the same turning effect as a larger force closer to the screw.

Worked example

B.5 Find the torque produced by a 20 N force on a 4 m-long rod free to rotate about an axis at its left end, for the three different force directions shown in Figure B.6.

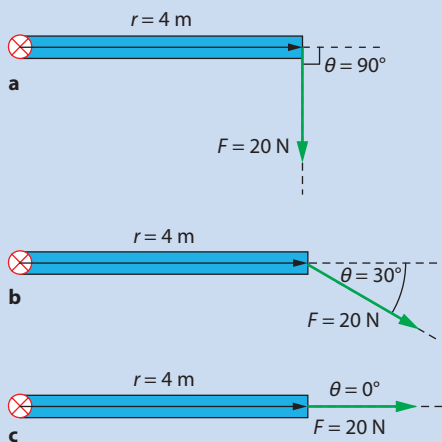


Figure B.6

a $\Gamma = Fr \sin \theta = 20 \times 4 \times \sin 90^\circ = 80 \text{ N m}$

b $\Gamma = Fr \sin \theta = 20 \times 4 \times \sin 30^\circ = 40 \text{ N m}$

c $\Gamma = Fr \sin \theta = 20 \times 4 \times \sin 0^\circ = 0 \text{ N m}$. This force is directed through the axis; it cannot turn the body and has zero torque.

There are other, equivalent ways to find torque, and you may choose whichever way you find convenient. In Figure B.7, for example, $r \sin \theta$ is the perpendicular distance d between the axis and the line of action of the force, so $\Gamma = Fd$.

Yet another way (Figure B.8) is to use components of the force. Noticing that $F \sin \theta$ is the component of the force along a direction at right angles to the vector r (PQ), we write $F_{\perp} = F \sin \theta$, so $\Gamma = F_{\perp} r$.

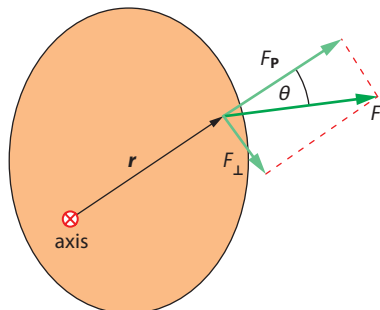


Figure B.8 Torque may be expressed as the product of the component of the force at right angles to the vector r and the distance between the axis and the point of application of the force.

Notice that if the force is directed through the axis, the torque is zero (Figure B.9). The force cannot produce a rotation in this case.

Torque is essential in discussing situations of equilibrium. For a point mass, equilibrium means that the net force on it is zero. For a rigid body, we have to distinguish between **translational** and **rotational equilibrium**. Translational equilibrium is similar to the equilibrium of a point mass: the net force on the body must be zero. The centre of mass of the rigid body remains at rest or moves in a straight line at constant velocity. Rotational equilibrium means that the net **torque** on the rigid body is zero. We examine these issues in Section B1.3.

B1.3 Equilibrium

Consider the simple example of a see-saw (Figure B.10). The plank can rotate about its pivot, the point of support. How can we find the force F_2 required for equilibrium? We do not want the plank to move or to rotate. We must apply the conditions for translational and rotational equilibrium.

Translational equilibrium requires that the net force be zero. In addition to the two forces F_1 and F_2 we have the upward reaction force N on the plank at the point of support. Translational equilibrium demands that,

$$F_1 + F_2 = N$$

For rotational equilibrium we demand that the net torque be zero. The reaction force N passes through the axis, so it produces no torque about the axis. Force F_1 tends to rotate the plank in a counter-clockwise direction, and force F_2 in a clockwise direction. The two forces have opposite torques, and we set them equal to one another to have zero net torque. This gives:

$$F_1 \times 1.2 = F_2 \times 0.8$$

$$240 \times 1.2 = F_2 \times 0.8$$

$$F_2 = 360 \text{ N}$$

From the translational-equilibrium condition, we then find that $N = 600 \text{ N}$.

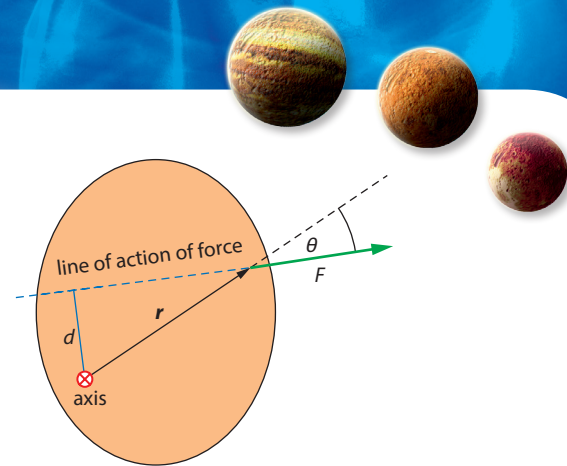


Figure B.7 Torque may be expressed as the product of the force and the perpendicular distance between the axis and the line of action of the force.

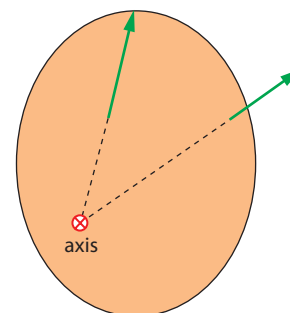


Figure B.9 Each of these forces passes through the axis and therefore produces zero torque about that axis.

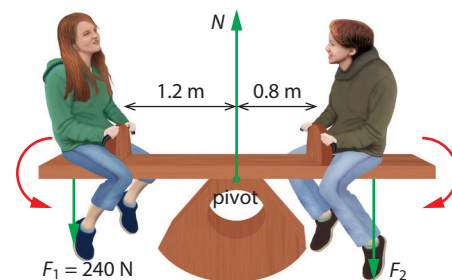


Figure B.10 A see-saw in equilibrium.

Worked example

B.6 Figure **B.11** shows a uniform ladder of length 5.0 m resting against a vertical wall, which is assumed to be frictionless. The other end rests on the floor, where a frictional force f prevents the ladder from slipping. The weight of the ladder is 350 N. Calculate the minimum coefficient of static friction between the ladder and the floor so that the ladder does not slip. The ladder is to make an angle of 60° with the floor.

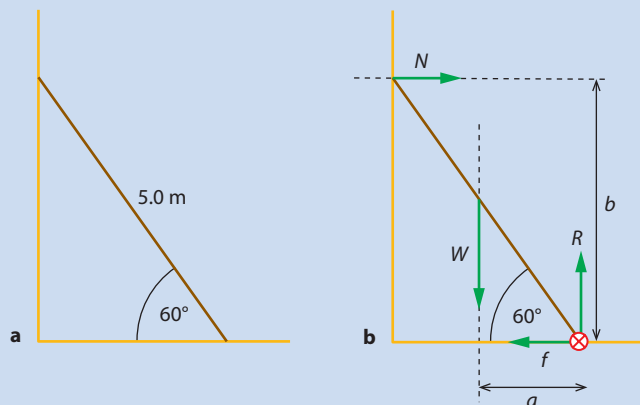


Figure B.11

Exam tip

- For translational equilibrium, require that the net force be zero
- For rotational equilibrium, require that the net torque be zero.
- You may take torques about any axis, not just the actual axis of rotation.

Figure **B.11a** shows the situation and Figure **B.11b** shows all the forces on the ladder.

Translational equilibrium demands that

$$N = f$$

$$R = W$$

where W is the weight of the ladder, R is the reaction force of the floor on the ladder and N is the normal force of the wall on the ladder. Since $W = 350$ N, we find $R = 350$ N right away.

We must now apply the condition for rotational equilibrium. But what is the axis of rotation? As stated in the Exam tip, we can take any point we wish as the rotation axis. It is convenient to choose a point through which as many forces as possible pass. In this way, their torques will be zero and we will not have to deal with them. Choosing the axis to pass through the point where the ladder touches the floor, we must then find the torques produced by W and N . The torque from N is $N \times b = N \times 5.0 \times \sin 60^\circ$. The torque from W is $W \times a = W \times 2.5 \times \cos 60^\circ$. The two torques must balance:

$$N \times 5.0 \times \sin 60^\circ = W \times 2.5 \times \cos 60^\circ$$

$$N = \frac{W \times 2.5 \times \cos 60^\circ}{5.0 \times \sin 60^\circ}$$

$$= 101 \text{ N}$$

This implies that $f = 101$ N. Now $f_{\max} = \mu_s R$, so $\mu_s = \frac{f_{\max}}{R} = \frac{101}{350} \approx 0.29$.

B1.4 Kinetic energy of a rotating body

Consider a rigid body that rotates with angular velocity ω about the axis shown in Figure B.12.

We break up the body into small bits, each of mass m_i . The kinetic energy of the rotating body is the sum of the kinetic energies of all the bits:

$$E_K = \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 + \frac{1}{2}m_3v_3^2 + \dots$$

$$= \sum \frac{1}{2}m_iv_i^2$$

Now, each bit is at a different distance r_i from the axis of rotation but all bits have the same angular velocity ω . Since $v_i = \omega r_i$ we deduce that:

$$E_K = \frac{1}{2}m_1\omega^2r_1^2 + \frac{1}{2}m_2\omega^2r_2^2 + \frac{1}{2}m_3\omega^2r_3^2 + \dots$$

$$= \frac{1}{2}\left(\sum m_ir_i^2\right)\omega^2$$

This is similar to the formula for the kinetic energy of a point mass, but here velocity is replaced by angular velocity and mass is replaced by the quantity $I = \sum m_ir_i^2$. We call this quantity the **moment of inertia** of the body about the rotation axis. Before continuing with kinetic energy, we must therefore discuss the idea of moment of inertia.

B1.5 Moment of inertia

Consider a point particle of mass m that is free to rotate about an axis a distance R from the particle (Figure B.13).

The moment of inertia of the particle about the given axis is:

$$I = \sum m_ir_i^2$$

But here there is just one 'bit' making up the body (the body itself), so the sum just has one term: $I = mR^2$. For two bodies of equal mass (Figure B.14), the moment of inertia is:

$$I = \sum m_ir_i^2 = mR^2 + mR^2 = 2mR^2$$

Notice that if we had chosen a different axis we would get a different moment of inertia.

Look now at a rigid body that is free to rotate about some axis. Consider a ring of radius R as in Figure B.15. We imagine that the ring consists of small bits of mass m_i . In this case, each bit has the same distance from the axis, so:

$$I = \sum m_ir_i^2 = \sum m_iR^2 = R^2\sum m_i = MR^2$$

where $M = \sum m_i$ is the total mass of the ring.

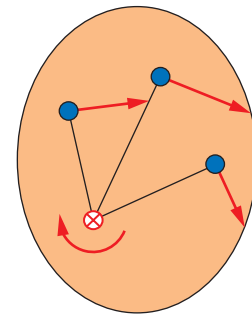


Figure B.12 The sum of the kinetic energies of each bit of the body equals the total kinetic energy of the entire body.

Exam tip

Moment of inertia is to rotational motion what mass is to linear motion.

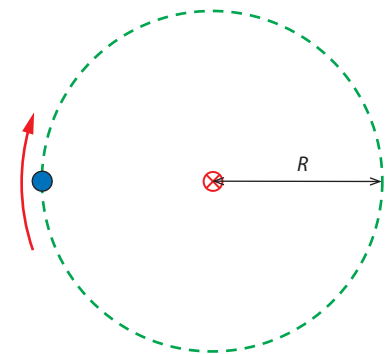


Figure B.13 A body rotating about a fixed axis.

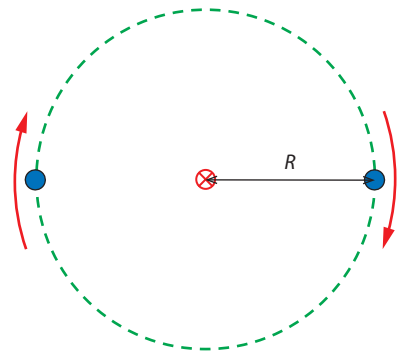


Figure B.14 Two bodies rotating about a common fixed axis.

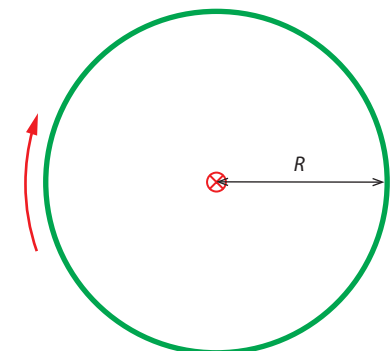


Figure B.15 A ring rotating about a fixed axis.

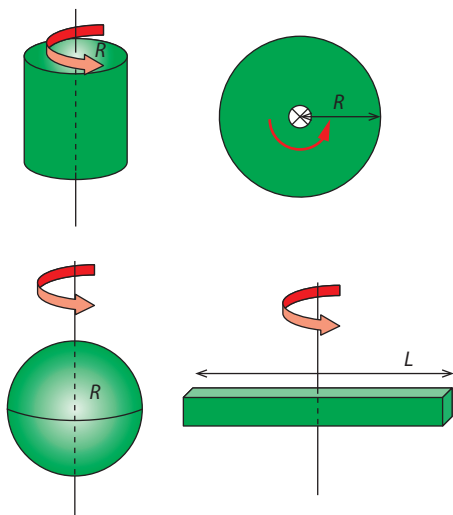


Figure B.16 Moments of inertia of a few common shapes, about the axes shown.

It is not possible to find the moments of inertia of other arrangements as easily as in the preceding cases. Here we just quote some results. The axis of rotation is indicated for each shape in Figure **B.16**.

Disc or cylinder: $I = \frac{1}{2}MR^2$

Sphere: $I = \frac{2}{5}MR^2$

Rod: $I = \frac{1}{12}ML^2$

We see that the moment of inertia is a quantity that depends on the mass of the body and how this mass is distributed around the rotation axis. The closer the mass is to the axis, the smaller the moment of inertia. The unit of moment of inertia is kgm^2 .

For a body rotating about some axis, its kinetic energy is then

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

where I is the body's moment of inertia about the rotation axis and ω is its angular velocity about the (same) rotation axis.

Worked example

B.7 A ring of mass 0.45 kg and radius 0.20 m rotates with angular velocity 5.2 rad s^{-1} . Calculate the kinetic energy of the ring.

The moment of inertia is $I = MR^2 = 0.45 \times 0.20^2 = 1.8 \times 10^{-2} \text{ kgm}^2$. The kinetic energy is therefore

$$E_K = \frac{1}{2}I\omega^2 = \frac{1}{2} \times 1.8 \times 10^{-2} \times (5.2)^2 = 0.24 \text{ J}$$

B1.6 Rolling without slipping

We will be dealing with bodies that do not just rotate about some axis but roll as well. For example, when you ride a bicycle the wheels rotate but at the same time they roll forward, making the bicycle move.

Consider a wheel of radius R that rotates about its axis (Figure **B.17a**). Every point in the body has the same angular velocity ω and the same

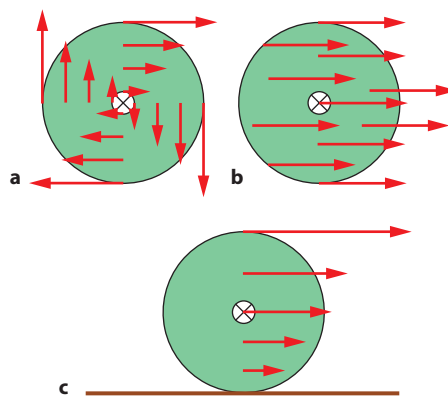
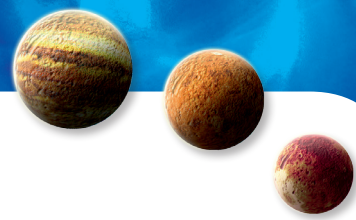


Figure B.17 **a** A rotating disc: each point of the disc has a different linear velocity. **b** A sliding disc: each point on the disc has the same velocity. **c** A disc that rolls without slipping: the point of contact has zero velocity.



linear speed $v = \omega R$. Now imagine the wheel sliding with velocity u along a horizontal floor but without rotating. In this case every point on the circumference has a velocity u to the right. Now let the wheel rotate as it slides to the right. Every point in the body now has two motions: due to rotation and due to sliding. Consequently, the top point has a velocity $v = \omega R + u$ to the right and the bottom point a velocity $v = \omega R - u$ to the left.

Rolling without slipping means that the point of contact between the wheel and the floor is instantaneously at rest. This means that $v = \omega R - u = 0$, so the velocity due to sliding, u , must be $u = \omega R$. So the bottom point is instantaneously at rest and the top point has velocity $v = \omega R + u = 2\omega R$ to the right.

This means that if we have a sphere of radius R and apply a force to it, the centre of mass of the sphere will start to slide in the direction of the force, and may also rotate about an axis. It all depends on where the force is applied. If the force is applied horizontally through the centre of mass, the sphere will slide but will not rotate. (There is no torque to cause rotation.) If the force is applied anywhere else it will cause sliding as well as rotation. But there is only one point where the sphere may be hit so that it slides without slipping. This is examined in a later section.

So sliding without slipping means that if the body rotates with angular velocity ω , the centre of mass of the body must have a velocity ωR , where R is the distance of the centre of mass from the point of contact.

Because angular acceleration is the rate of change of angular velocity, the condition of rolling without slipping may be expressed in terms of angular acceleration as $a = \alpha r$.

B1.7 Kinetic energy of a body that rolls without slipping

We saw that a rotating body has kinetic energy $E_K = \frac{1}{2}I\omega^2$. If the body rolls in addition, then we have to include the kinetic energy due to the translational motion of the centre of mass. The total kinetic energy is then $E_K = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2$, where v is the speed of the centre of mass.

If the centre of mass of the body is at a height h from some horizontal level then the gravitational potential energy relative to that level is Mgh , so the total mechanical energy is:

$$E_T = \frac{1}{2}I\omega^2 + \frac{1}{2}Mv^2 + Mgh$$

In the absence of resistance forces, the total mechanical energy is conserved.

Worked example

B.8 A cylinder of mass M and radius R (moment of inertia $I = \frac{1}{2}MR^2$) begins to roll from rest down an inclined plane (Figure B.18). Calculate the linear speed of the cylinder when it reaches level ground a height h lower.

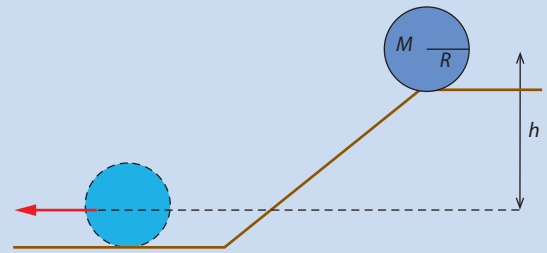


Figure B.18

The initial energy, at the upper level, is Mgh . The energy at the lower level is $E_K = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2$. Equating the two energies gives:

$$\frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2 = Mgh$$

The moment of inertia of the cylinder is $I = \frac{1}{2}MR^2$. Because the cylinder rolls without slipping, $v = \omega R$, or $\omega = \frac{v}{R}$. Therefore,

$$\frac{1}{2}Mv^2 + \frac{1}{2}\left(\frac{1}{2}MR^2\right)\frac{v^2}{R^2} = Mgh$$

$$\frac{1}{2}v^2 + \frac{1}{4}v^2 = gh$$

$$\frac{3v^2}{4} = gh$$

The linear speed is $\sqrt{\frac{4gh}{3}}$

Exam tip

Notice that if the body were a point mass the answer for the speed would be $v = \sqrt{2gh}$, greater than the answer here.

Exam tip

Moment of inertia is to rotational motion what mass is to linear motion.

Torque is to rotational motion what force is to linear motion.

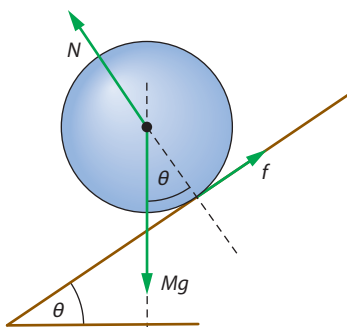


Figure B.19 A sphere that rolls without slipping down an inclined plane: there must be a frictional force to provide the torque for it to turn.

B1.8 Newton's second law applied to rotational motion

When we solve problems in mechanics with rigid bodies we must apply Newton's second law to the centre of mass of the body:

$$F_{\text{net}} = Ma \quad (\text{Newton's second law for translational motion})$$

that is, in exactly the way we would treat a point mass. But we must also apply Newton's second law to the rotational motion of the body; it can be shown that this law takes the form

$$\Gamma_{\text{net}} = I\alpha \quad (\text{Newton's second law for rotational motion})$$

(This is to be expected since torque plays the role of force and moment of inertia plays the role of mass.)

We begin with an example of a sphere rolling without slipping down an inclined plane that makes an angle θ with the horizontal (Figure B.19). The sphere has mass M and radius R . We want to find the linear acceleration of the sphere as it comes down the plane. The diagram shows the forces on the sphere. Notice right away that a frictional force f must be present. Without one, the sphere would just slide down the plane and would not roll. (The problem would then be identical to that for a point mass.)



The net force on the sphere along the incline is $Mg\sin\theta - f$, so:

$$F_{\text{net}} = Ma$$

$$Mg\sin\theta - f = Ma \quad (\text{Newton's second law for translational motion})$$

Because the forces perpendicular to the incline are in balance, we also have that $N = Mg\cos\theta$.

The net torque about a horizontal axis through the centre of mass is fR (N and Mg have zero torques about this point) and so, because the moment of inertia of a sphere about its axis is $I = \frac{2}{5}MR^2$:

$$\Gamma = I\alpha$$

$$fR = \frac{2}{5}MR^2\alpha \quad (\text{Newton's second law for rotational motion})$$

We assume the sphere is rolling without slipping, so

$$\alpha = \frac{a}{R}$$

So our two equations become:

$$Mg\sin\theta - f = Ma$$

$$fR = \frac{2}{5}MR^2\frac{a}{R}$$

These simplify to:

$$Mg\sin\theta - f = Ma$$

$$f = \frac{2}{5}Ma$$

and so:

$$Mg\sin\theta - \frac{2}{5}Ma = Ma$$

giving finally

$$a = \frac{5g\sin\theta}{7}$$

Let us now calculate the speed of the sphere when its centre of mass is lowered by a vertical distance h . The distance travelled down the plane is s and, from Figure B.20, $h = s\sin\theta$. So:

$$v^2 = 2as = 2 \times \frac{5g\sin\theta}{7} \times \frac{10gh}{7}$$

We get the same result if we apply conservation of energy:

$$\frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2 = Mgh$$

$$\frac{1}{2}Mv^2 + \frac{1}{2}\left(\frac{2}{5}MR^2\right)\frac{v^2}{R^2} = Mgh$$

$$\frac{1}{2}v^2 + \frac{1}{5}v^2 = gh$$

$$\frac{7v^2}{10} = gh$$

$$v^2 = \frac{10gh}{7}$$



The power of formal analogies

Once the correspondence between linear and angular quantities is established, the formulas relating the angular quantities can be guessed from the corresponding formulas for linear quantities. The formalism is the same and so the formulas are the same. We have seen similar connections between electricity and gravitation, to name one example, before.

Exam tip

Notice that if the body were a point mass, the acceleration would be just $a = g\sin\theta$.

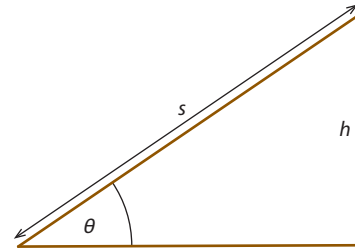


Figure B.20 Relation between the vertical distance and the distance along the plane.

At this point you may be wondering how we can claim energy conservation when there is a frictional force present! The point is that the point of contact where the frictional force acts is not sliding; it is a point that is instantaneously at rest and so the frictional force does not transfer any energy into thermal energy. It would do so if we had rolling with slipping, but we will not consider such cases in this course.

Worked examples

B.9 A block of mass m is attached to a string that goes over a cylindrical pulley of mass M and radius R (Figure B.21). When the block is released, the pulley begins to turn as the block falls. Calculate the acceleration of the block.

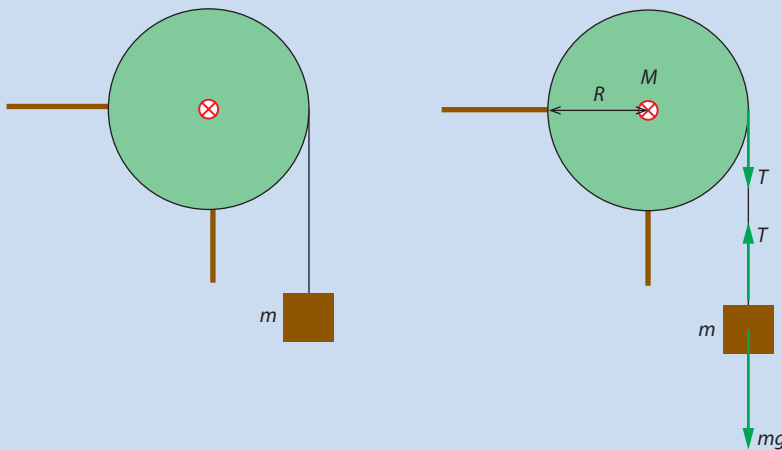


Figure B.21

Applying the second law to the motion of the hanging mass,

$$mg - T = ma$$

Applying the second law to the rotation of the pulley,

$$TR = I\alpha$$

The angular and linear accelerations are related by $a = \alpha R$. So the two equations become

$$TR = I \frac{a}{R}$$

$$mg - T = ma$$

Solving the first equation for T , $T = \frac{Ia}{R^2} = \frac{1}{2}MR^2 \frac{a}{R^2} = \frac{1}{2}Ma$. Substitute this in the second equation to get

$$mg - \frac{1}{2}Ma = ma$$

$$mg = a \left(m + \frac{M}{2} \right)$$

$$a = \frac{mg}{m + \frac{M}{2}}$$



B.10 A snooker ball of mass M and radius R is hit horizontally at a point a distance d above its centre of mass (Figure B.22). Determine d such that the ball rolls without slipping.

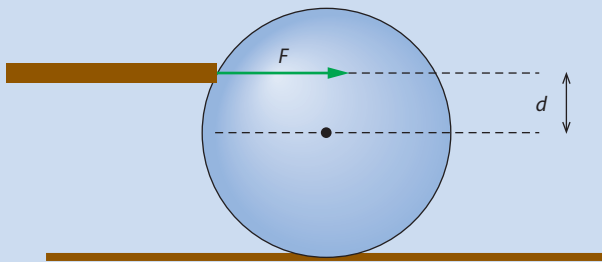


Figure B.22

The torque on the ball about an axis through the centre of mass is Fd . Therefore

$$Fd = \frac{2}{5}MR^2\alpha$$

The net force on the ball is F , so

$$F = Ma$$

The ball does not slip, and so the angular and linear accelerations are related by $a = \alpha R$, so

$$Fd = \frac{2}{5}MR^2 \frac{a}{R}$$

$$F = Ma$$

Dividing one equation by the other to get rid of F , we get $d = \frac{2}{5}R$.

B1.9 Work and power

Consider a constant force F applied to a rigid body for a time Δt . Let this force move its point of application by a distance Δs , rotating the body by an angle $\Delta\theta$ in the process. The force does work $W = F\Delta s$. But $\Delta s = R\Delta\theta$, so the work is also given by:

$$W = FR\Delta\theta$$

The torque of the force is $\Gamma = FR$, so

$$W = \Gamma\Delta\theta.$$

The familiar result from linear mechanics that the work done by a net force equals the change in kinetic energy holds here as well, for the work done by a net torque.

The power developed by the force is:

$$P = \frac{W}{\Delta t} = \frac{\Gamma\Delta\theta}{\Delta t}$$

$$P = \Gamma\omega$$

Both formulas are direct analogues of the corresponding quantities in linear motion (Table B.1).

Linear motion	Rotational motion
$F = ma$	$\Gamma = I\alpha$
$W = F\Delta s$	$W = \Gamma\Delta\theta$
$P = Fv$	$P = \Gamma\omega$

Table B.1 Force, work and power for linear and angular quantities.

Worked example

B.11 A disc of moment of inertia 25 kgm^2 and radius 0.80 m which is rotating at 320 revolutions per minute must be stopped in 12 s . Calculate **a** the work needed to stop it, and **b** the average power developed in stopping it. **c** Determine the torque that stopped the disc (assuming it is constant).

a The initial kinetic energy of the rotating disc is $E_K = \frac{1}{2}I\omega^2$. We must find ω : $\omega = \frac{320 \times 2\pi}{60} = 33.5 \text{ rad s}^{-1}$.
Hence $E_K = \frac{1}{2} \times 25 \times 33.5^2 = 1.4 \times 10^4 \text{ J}$. This is the work required.

b The average power is therefore $P = \frac{1.4 \times 10^4}{12} = 1.2 \times 10^3 \text{ W}$.

c We will use $P = \Gamma\omega$. For the average power we use the average angular speed, so $\bar{P} = \Gamma\bar{\omega}$.
Since $\bar{\omega} = \frac{33.5}{2} = 16.75 \text{ rad s}^{-1}$, the torque stopping the disc is $\Gamma = \frac{1.2 \times 10^3}{16.75} \approx 72 \text{ Nm}$.

B1.10 Angular momentum

The angular momentum of a rigid body with moment of inertia I rotating about a fixed axis with angular speed ω is defined as:

$$L = I\omega$$

(See Figure **B.23**.) If the body is a point mass, the angular momentum is $L = I\omega = mr^2\omega = m(\omega r)r = mvr$, an expression we used in Topic **12** in relation to the Bohr model. The unit of angular momentum is kg m s^{-2} (equivalent to Js).

Just as there is a relation between linear momentum and net force in particle mechanics, here there is a relation between angular momentum and net torque:

$$\Gamma_{\text{net}} = \frac{\Delta L}{\Delta t}$$

The net torque on a body is the rate of change of the angular momentum of the body.

Similarly, we have the law of **conservation of angular momentum**:

When the net torque on a system is zero, the angular momentum is conserved, that is, it stays constant.

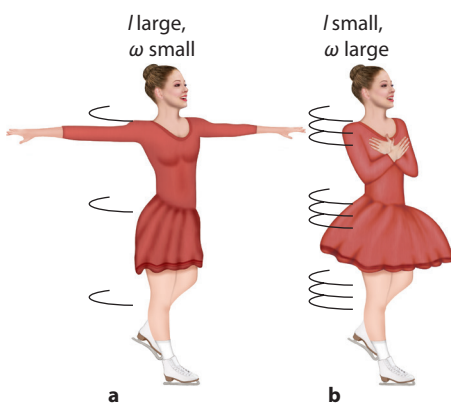


Figure B.23 **a** A skater rotates about a vertical axis. **b** When she brings her arms in, her moment of inertia is reduced and her angular velocity increases in order to conserve angular momentum.



Worked examples

B.12 A spherical star of mass M and radius R rotates about its axis with angular speed ω . The star explodes, ejecting mass into space. The star that is left behind has mass $\frac{M}{10}$ and radius $\frac{R}{20}$. Calculate the new angular speed of the star.

Angular momentum is conserved, so

$$\frac{2}{5}MR^2\omega = \frac{2}{5}\frac{M}{10}\left(\frac{R}{20}\right)^2\omega' \Rightarrow \omega' = 10 \times 20^2\omega = 4000\omega$$

B.13 A horizontal disc of mass $M=0.95$ kg and radius $R=0.35$ m rotates about a vertical axis with angular speed 3.6 rad s⁻¹. A piece of modelling clay of mass $m=0.30$ kg lands vertically on the disc, attaching itself at a point $r=0.25$ m from the centre. Find the new angular speed of the disc.

The original angular momentum of the disc is

$$L_{\text{in}} = I\omega = \frac{1}{2}MR^2\omega = 0.209 \text{ J s}$$

The final angular momentum is

$$L_{\text{fin}} = I\omega' + mr^2\omega' = \left(\frac{1}{2}2.2 \times 0.35^2 + 0.12 \times 0.25^2\right)\omega'$$

Angular momentum will be conserved because there are no external torques on the disc–clay system, so

$$\left(\frac{1}{2}0.95 \times 0.35^2 + 0.30 \times 0.25^2\right)\omega' = 0.209$$
$$\omega' = 2.7 \text{ rad s}^{-1}$$

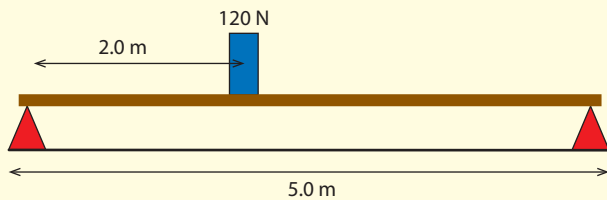
Nature of science

Adapting models to the real world

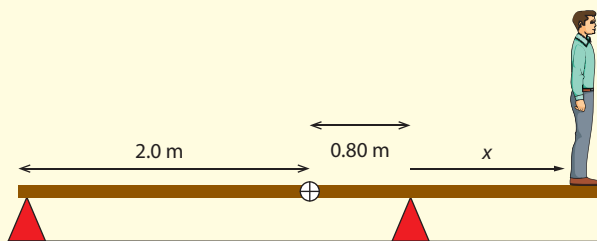
The point particle model simplifies interactions between objects, but does not predict how real objects behave. By adapting models to take rotation into account, engineers can apply the laws of physics to design structures and machines that aim to make our lives safe, comfortable and enjoyable. The great technological advances in energy production, food production, transportation, telecommunications, space exploration, weather prediction, medicine, the entertainment industry and a myriad of other fields are all the result of the application of the basic laws and principles of the sciences to practical problems.

? Test yourself

- 1 A disc has an initial angular velocity 3.5 rad s^{-1} and after 5.0 s the angular velocity increases to 15 rad s^{-1} . Determine the angle through which the disc has turned during that 5.0 s .
- 2 A body rotates about an axis with an angular velocity of 5.0 rad s^{-1} . The angular acceleration is 2.5 rad s^{-2} . Calculate the body's angular velocity after it has turned through an angle of 54 rad .
- 3 A body rotates with an initial angular velocity of 3.2 rad s^{-1} . The angular velocity increases to 12.4 rad s^{-1} in the course of 20 full revolutions. Calculate the angular acceleration.
- 4 A uniform plank of weight 450 N and length 5.0 m is supported at both ends. A block of weight 120 N is placed at a distance of 2.0 m from the left end. Calculate the force at each support.

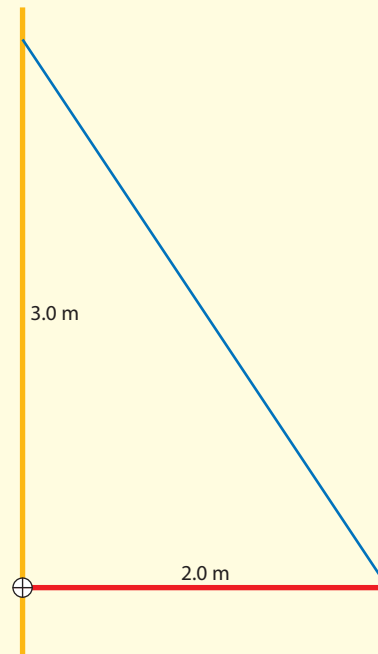


- 5 A uniform plank of mass 30 kg and length 4.0 m is supported at its left end and at a point 0.80 m from the middle.



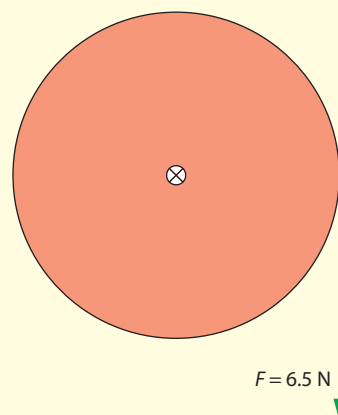
Calculate the largest distance x to which a boy of mass 40 kg can walk without tipping the rod over.

- 6 A uniform plank of length 2.0 m and weight 58 N is supported horizontally by a cable attached to a vertical wall.



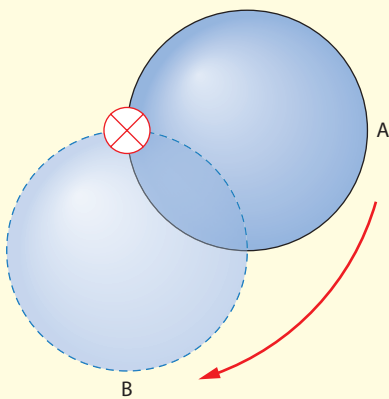
Calculate **a** the tension in the cable, and **b** the magnitude and direction of the force exerted by the wall on the rod.

- 7 A cylinder of mass 5.0 kg and radius 0.20 m is attached to an axle parallel to its axis and through its centre of mass. A constant force of 6.5 N acts on the cylinder as shown in below. Find the angular speed of the cylinder after 5.0 s .



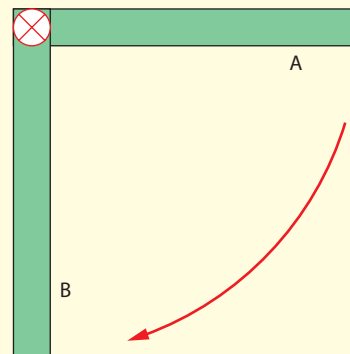


- 8 A point mass, a sphere, a cylinder and a ring each have mass M . The solid bodies have the same radius R . The four bodies are released from the same position on an inclined plane (with the ring upright, so it rolls). Determine the order, from least to greatest, of their speeds when they reach level ground. Assume rolling without slipping.
- 9 A disc of mass 12 kg and radius 0.35 m is spinning with angular velocity 45 rad s^{-1} .
- Determine how many revolutions per minute (rpm) the disc is making.
 - A force is applied to the rim of the disc, tangential to the disc's circumference. Determine the magnitude of this force if the disc is to stop spinning after 4.0 s .
 - Calculate the number of revolutions it makes before stopping.
- 10 A uniform sphere rotates about a fixed horizontal axis through the edge of the sphere, as shown below. The axis is at the height of the initial position of the sphere's centre of mass. The moment of inertia of the sphere about this axis is $I = \frac{7}{5}MR^2$. The sphere is released from rest at position A.



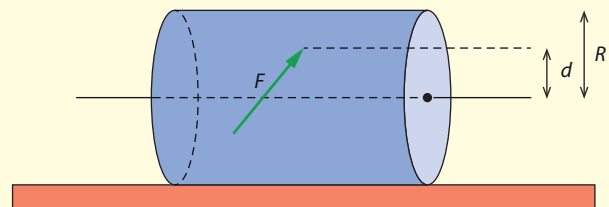
- Find an expression for the initial angular acceleration of the sphere.
- State and explain whether the angular acceleration is constant in magnitude as the sphere rotates.
- Find an expression for the angular velocity of the sphere as it moves past position B.

- 11 A rod of length $L = 1.20\text{ m}$ and mass $M = 3.00\text{ kg}$ is free to move about a fixed axis at its left end. Its moment of inertia about this axis is given by $\frac{1}{3}ML^2$.



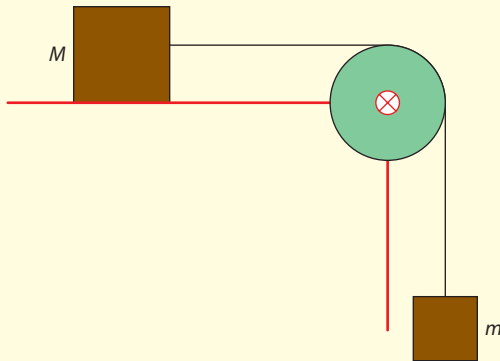
The rod is released from rest in the horizontal position A.

- Calculate the initial angular acceleration of the rod.
 - State and explain whether the angular acceleration is constant in magnitude as the rod rotates.
 - Calculate the angular velocity of the rod as it moves past the vertical position B.
- 12 A horizontal force F is applied to the surface of a cylinder of mass M and radius R . The force is applied a vertical distance d above its centre of mass, as shown below.

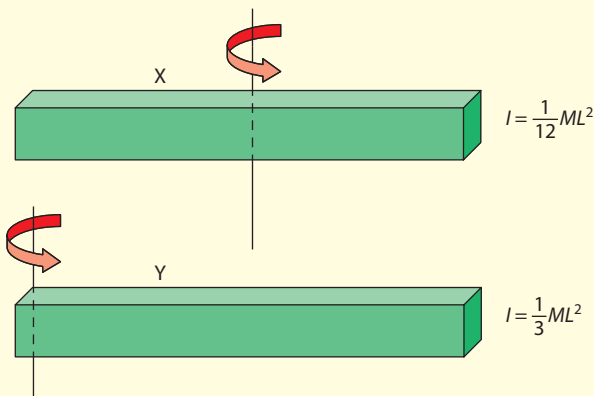


Determine d as a fraction of R such that the cylinder rolls without slipping. The moment of inertia of a cylinder about its central axis is $\frac{1}{2}MR^2$.

- 13 A block of mass m hangs from the end of a string that goes over a pulley of radius R , and is connected to another block of mass M that rests on a horizontal table.

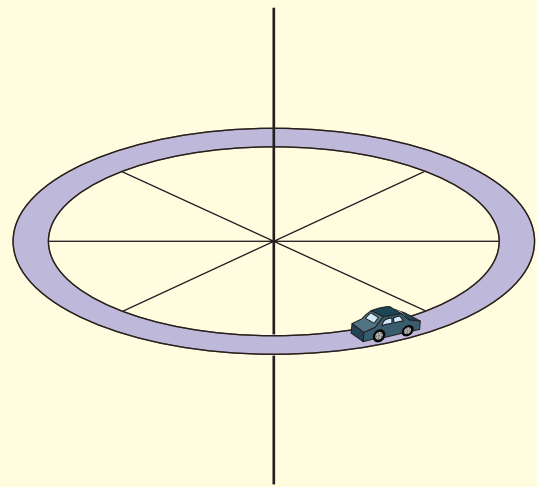


- a Assuming that the pulley is massless and that there are no frictional forces, calculate the acceleration of each block when the smaller block is released.
- b The pulley is now assumed to have mass M . The small block is again released and the pulley turns as the blocks move. Again calculate the acceleration of each block. (*Hint: the tensions in the two strings are different.*)
- 14 A horizontal disc, with radius 0.80 m and moment of inertia 280 kgm^2 about its vertical axis, rotates about this axis at 320 revolutions per minute. The disc is brought to rest in 12 s . Calculate **a** the work done, and **b** the power developed in stopping the disc.
- 15 Two identical rods, X and Y, each of length 1.20 m and mass 2.40 kg , are made to rotate about different vertical axes, as shown below each with angular velocity 4.50 rads^{-1} .



Calculate **a** the kinetic energy and **b** the angular momentum of X and of Y.

- 16 A star of mass M and radius R explodes radially and symmetrically. The star is left with a mass of $\frac{1}{10}M$ and a radius of $\frac{1}{50}R$. Calculate the ratio of the star's final angular velocity to its initial angular velocity.
- 17 A battery-driven toy car of mass 0.18 kg is placed on a circular track that is part of a horizontal ring with radius 0.50 m and moment of inertia 0.20 kgm^2 relative to its vertical axis. The ring can rotate about this axis without friction.



The car is started and its speed is measured to be 0.80 ms^{-1} relative to the ground. Calculate the angular speed of the ring.



B2 Thermodynamics

Thermodynamics deals with the conditions under which heat can be transformed into mechanical work. The **first law of thermodynamics** states that the **thermal energy** given to a system is used to increase the system's internal energy and to do mechanical work. The **second law of thermodynamics** invokes limitations on how much thermal energy can actually be transformed into work.

B2.1 Internal energy

In Topic 3 we defined the internal energy of a gas as the total random kinetic energy of the particles of the gas plus the potential energy associated with intermolecular forces. If the gas is assumed to be ideal, the intermolecular forces are strictly zero and the internal energy of the gas is just the total random kinetic energy of the particles of the gas. We have seen that the average kinetic energy of the particles is given by:

$$\overline{E_K} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

where $k = \frac{R}{N_A} = 1.38 \times 10^{-23} \text{ J K}^{-1}$, is the Boltzmann constant.

The internal energy U of an ideal gas with N particles is $N\overline{E_K}$, so

$$U = \frac{3}{2}NkT$$

or, since $k = \frac{R}{N_A}$ and $PV = nRT$,

$$U = \frac{3}{2}nRT = \frac{3}{2}PV$$

where n is the number of moles.

This formula shows that the internal energy of a fixed number of moles of an ideal gas depends only on its temperature and not on the nature of the gas, its volume or other variables.

The **change** in internal energy due to a change in temperature is thus:

$$\Delta U = \frac{3}{2}nR\Delta T$$

Learning objectives

- Understand the first and second laws of thermodynamics.
- Appreciate the concept of entropy.
- Work with cyclic processes.
- Identify and understand isovolumetric, isobaric, isothermal and adiabatic processes.
- Understand the Carnot cycle and thermal efficiency.

Exam tip

You must know the equivalent ways of expressing internal energy.

Exam tip

Throughout this section, we will deal with monatomic gases.

Worked example

B.14 A flask contains a gas at a temperature of 300 K. The flask is taken aboard a fast-moving aircraft. Suggest whether the temperature of the gas will increase as a result of the molecules now moving faster. The temperature inside the aircraft is 300 K and the container is not insulated.

No. The temperature of the gas depends on the random motion of the molecules and not on any additional uniform motion imposed on the gas as a result of the motion of the container.

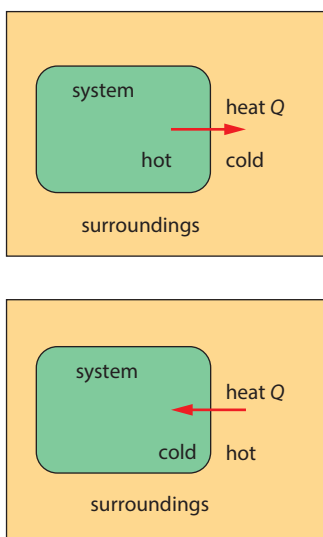


Figure B.24 A system and its surroundings.

Process	Description
Isothermal	Temperature stays constant
Isobaric	Pressure stays constant
Isovolumetric	Volume stays constant
Adiabatic	No heat enters or leaves the system

Table B.2 A few thermodynamic processes.

B2.2 Systems and processes

In thermodynamics we often deal with **systems**; this means the complete set of objects under consideration. Thus, a gas in a container is a system, as is a certain mass of ice in a glass. What is not in the system is the **surroundings** of the system. Heat can enter or leave a system depending on its temperature relative to that of the surroundings (Figure B.32).

A system can be **open** or **closed**: mass can enter and leave an open system but not a closed system. An **isolated** system is one in which no energy in any form enters or leaves. If all the parameters defining the system are given, we speak of the system being in a particular **state**. For example, the state of a fixed quantity of an ideal gas is specified if its pressure, volume and temperature are specified. Any processes that change the state of a system are called **thermodynamic processes**. Thus, heating a gas may result in a changed pressure, temperature or volume and is thus a thermodynamic process. Doing work on the gas by compressing it is another thermodynamic process.

Internal energy is a property of the particular state of the system under consideration, and for this reason internal energy is called a **state function**. Thus, if two equal quantities of an ideal gas originally in different states are brought to the same state (i.e. same pressure, volume and temperature), they will have the same internal energy, irrespective of the original state and how the gas was brought to the final state. By contrast, heat and work are not state functions. We cannot speak of the heat content of a system or of its work content. Heat and work are related to **changes** in the state of the system, not to the state itself.

In this course we will be mainly interested in four different types of thermodynamic processes. These are defined in Table B.2.

B2.3 Pressure–volume diagrams

Changes in a gas may be conveniently shown on a pressure–volume diagram. We saw examples of this in Section 3.2. Examples of **isovolumetric** (red) and **isobaric** (blue) processes are shown in Figure B.25.

A process that we did not examine in Section 3.2 is the **adiabatic** process. In an adiabatic process, no heat enters or leaves the system. It can be shown that, if the state of a fixed quantity of an ideal gas is changed from pressure p_1 and volume V_1 to p_2 and V_2 in an adiabatic process,

$$p_1 V_1^{5/3} = p_2 V_2^{5/3}$$

$$p V^{5/3} = \text{constant} = c_1$$

This will mean that, for an adiabatic changes, the temperature also changes.

From the ideal gas law we have that $\frac{pV}{T} = \text{constant} = c_2$. Solving this for the pressure gives $p = \frac{c_2 T}{V}$. Substituting this into the adiabatic law,

$$\left(\frac{c_2 T}{V}\right) V^{5/3} = c_1$$

$$T V^{2/3} = \text{constant} \text{ or } T_1 V_1^{2/3} = T_2 V_2^{2/3}$$

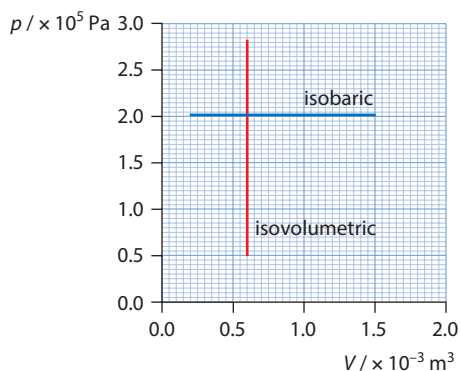
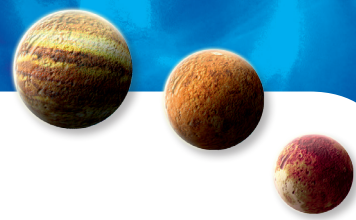


Figure B.25 Isobaric and isovolumetric processes.



So if a gas expands adiabatically, the temperature drops. This sounds surprising the first time you hear it: the expansion is adiabatic, no heat enters or leaves, so why should the temperature change? We will be able to explain this as soon as we learn about the first law of thermodynamics.

Worked example

B.15 An ideal gas expands adiabatically from a state with pressure 5.00×10^5 Pa, volume $2.20 \times 10^{-3} \text{ m}^3$ and temperature 485 K to a new volume of $3.80 \times 10^{-3} \text{ m}^3$. Calculate the new pressure and new temperature of the gas.

From $p_1 V_1^{5/3} = p_2 V_2^{5/3}$ we find:

$$5.00 \times 10^5 \times (2.20 \times 10^{-3})^{5/3} = p_2 \times (3.80 \times 10^{-3})^{5/3}$$

$$p_2 = \frac{5.00 \times 10^5 \times (2.20 \times 10^{-3})^{5/3}}{(3.80 \times 10^{-3})^{5/3}}$$

$$= 2.01 \times 10^5 \text{ Pa}$$

To find the new temperature we may use $T_1 V_1^{2/3} = T_2 V_2^{2/3}$, but it is simpler to use the ideal gas law and write $\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2}$, so

$$\frac{5.00 \times 10^5 \times 2.20 \times 10^{-3}}{485} = \frac{2.01 \times 10^5 \times 3.80 \times 10^{-3}}{T_2}$$

$$T_2 = \frac{2.01 \times 10^5 \times 3.80 \times 10^{-3} \times 485}{5.00 \times 10^5 \times 2.20 \times 10^{-3}}$$

$$= 337 \text{ K}$$

An adiabatic expansion on a pressure–volume diagram looks like an **isothermal** expansion but is steeper. Figure **B.26a** shows an isothermal (blue) and an adiabatic (red) *expansion* of the same gas from a common state. Figure **B.26b** shows an isothermal and adiabatic *compression* of the same gas from a common state.

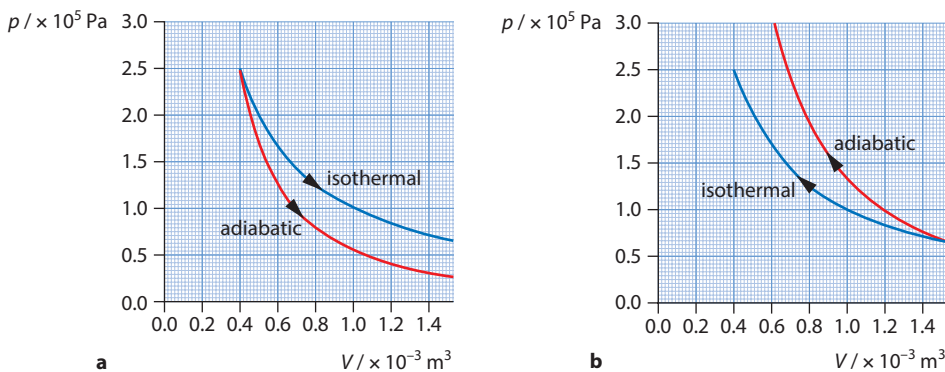


Figure B.26 Isothermal (blue) and adiabatic (red) changes of state: **a** expansion from identical initial states; **b** compression from identical initial states. In both cases the adiabatic curve is steeper.

B2.4 Work done on or by a gas

Consider a given quantity of a gas in a container with a frictionless, movable piston (Figure B.27). If the piston is to stay in place, a force from the outside has to be applied to counterbalance the force due to the pressure of the gas. Let us now compress the gas by pushing the piston in, with a very slightly greater force.

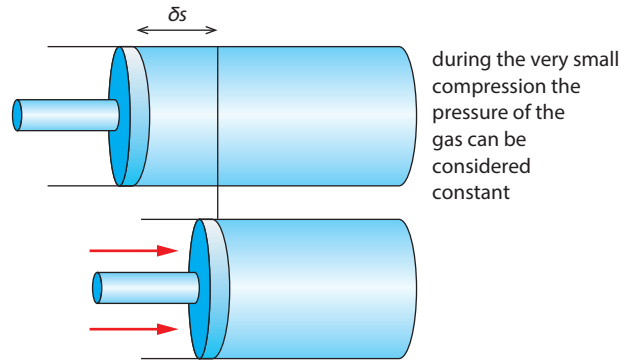


Figure B.27 When the piston is pushed in by a small amount, work is done on the gas.

If the initial pressure of the gas is p and the cross-sectional area of the piston is A , then the force with which one must push is pA . If the pressure stays constant, the force is constant and so we may use the result from mechanics that work is force times distance moved in the direction of the force. The piston moves a distance Δs , so the work done is:

$$W = F\Delta s$$

$$= pA\Delta s$$

But $A\Delta s$ is change in the volume of the gas, ΔV . The same result holds if, instead, the piston is moved outwards by the gas as it expands.

The work done when the volume of a gas is changed by ΔV at constant pressure is $W = p\Delta V$.

Figure B.28 shows that the work done is the area under the isobaric curve. The work done in this case is:

$$W = 1.5 \times 10^5 \times (1.6 \times 10^{-3} - 0.4 \times 10^{-3}) = 180 \text{ J}$$

What if the pressure changes? Then the force is not constant and we have to do what we did in mechanics: the work done is the area under the curve in a force–distance graph (Figure B.29). Here the corresponding result is the area under the pressure–volume graph.

The work done when a gas expands by an arbitrary amount is the area under the curve in the pressure–volume diagram.

In Figure B.30, we must calculate the area of the shaded trapezoid:

$$W = \frac{(1.4 + 2.6) \times 10^5}{2} \times 1.2 \times 10^{-3} = 180 \text{ J}$$

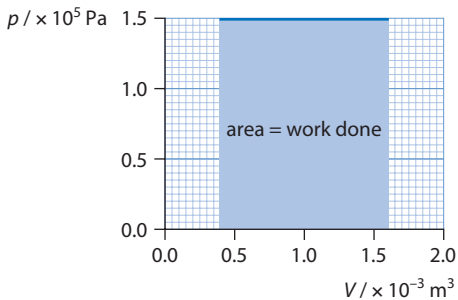


Figure B.28 The area under the graph in a pressure–volume diagram is equal to the work done.

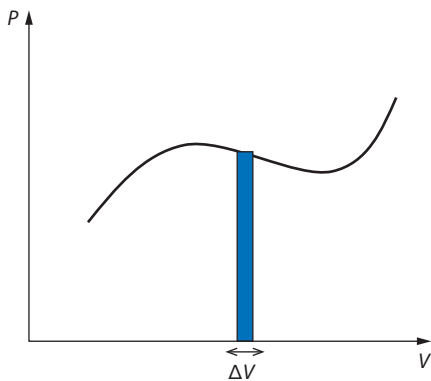


Figure B.29 Although the pressure is not constant, the work done can be calculated for a series of infinitesimal volume changes.

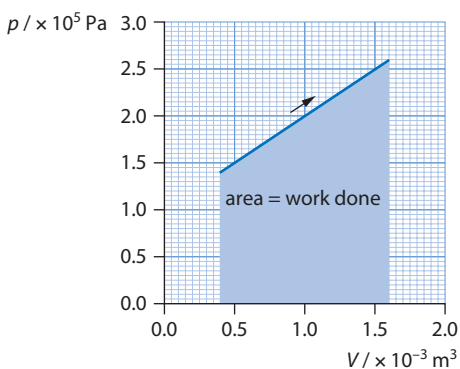
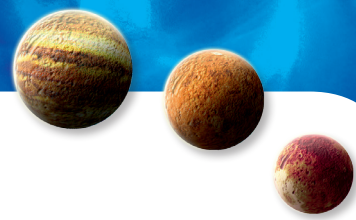


Figure B.30 The work done is found from the area under the curve in the pressure–volume diagram.



The pressure–volume diagrams in Figure **B.31** show an ideal gas that expands from state A to state B and is then compressed back to state A. The gas does work in expanding from A to B but work is done **on** the gas when it is compressed back to A. The net work done is therefore the area of the loop.

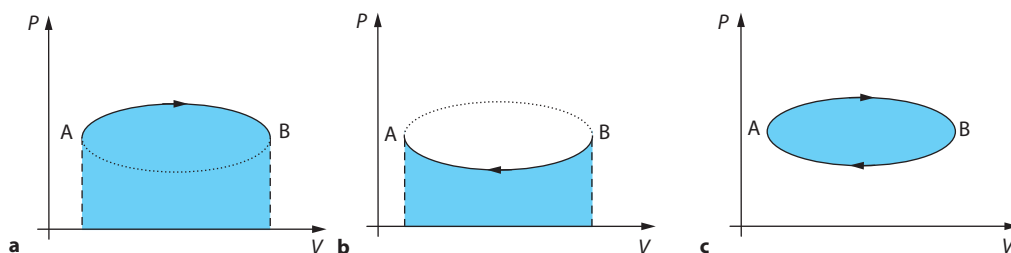


Figure B.31 For a closed loop in a pressure–volume diagram, the work done is the area within the loop.

For a closed loop in a pressure–volume diagram, the net work done is the area within the loop.

Worked example

B.16 A gas is compressed at a constant pressure of $2.00 \times 10^5 \text{ Pa}$ from a volume of 2.00 m^3 to a volume of 0.500 m^3 . The temperature is initially 40.0°C .

- Find the work done.
- Calculate the final temperature of the gas.

a Since the compression takes place under constant pressure, the work done is

$$p \times (\text{change in volume}) = 2.00 \times 10^5 \text{ Pa} \times 1.50 \text{ m}^3 \\ = 3.00 \times 10^5 \text{ J}$$

b The final temperature is found using $\frac{V}{T} = \text{constant}$, so $\frac{2.00}{313} = \frac{0.500}{T}$, giving $T = 78.25 \text{ K} = -195^\circ\text{C}$.

B2.5 The first law of thermodynamics

An amount of heat Q given to a gas will increase the internal energy of the gas and/or will do work by expanding the gas. Conservation of energy demands that:

$$Q = \Delta U + W$$

where ΔU is the change in internal energy and W is the work done. This formula goes with certain conventions that you must know:

- $Q > 0$ means heat is supplied to the gas
- $\Delta U > 0$ means the internal energy of the gas increases
- $W > 0$ means that work is done by the gas as it expands.

Conversely,

- $Q < 0$ means heat is removed from the gas
- $\Delta U < 0$ means the internal energy of the gas decreases
- $W < 0$ means that work is done on the gas by compressing it.

Exam tip

The conventions for signs in the first law are crucial.

Worked examples

B.17 An ideal gas in a container with a piston expands isothermally. Energy $Q = 2.0 \times 10^5 \text{ J}$ is transferred to the gas. Calculate the work done by the gas.

The process is isothermal, so $T = \text{constant}$. It follows that $\Delta U = 0$, and since $Q = \Delta U + W$ we must have $W = Q$. So, the work done by the gas in this case is equal to the energy supplied to it: $2.0 \times 10^5 \text{ J}$.

B.18 An ideal gas expands adiabatically.

- Explain why the temperature decreases.
- Use your answer to **a** to explain why the adiabatic curve of an ideal gas, expanding from a given state, is steeper than the corresponding isothermal curve from the same state.

- a** We have $Q = 0$ because the process is adiabatic. The first law, $Q = \Delta U + W$, then gives $0 = \Delta U + W$, from which we find $\Delta U = -W$. The gas is expanding, so it is doing work: $W > 0$. Therefore $\Delta U < 0$, that is, the internal energy and thus the temperature decrease. Similarly, if the gas were compressed adiabatically, the temperature would increase.
- b** Consider an adiabatic and an isothermal process, both starting from the same point and bringing the gas to the same (expanded) final volume. Since the adiabatic process reduces the temperature and the isothermal process does not, the final pressure after the adiabatic expansion will be lower than the final pressure after the isothermal expansion. Hence, the adiabatic curve must be steeper (see Figure **B.26a**).

B.19 An ideal gas, kept at a constant pressure of $3.00 \times 10^6 \text{ Pa}$, has an initial volume of 0.100 m^3 . The gas is compressed at constant pressure down to a volume of 0.080 m^3 . Find **a** the work done on the gas and **b** the energy transferred.

- a** The work done is $W = p \times \Delta V = 3.00 \times 10^6 \times (-0.020) = -6.00 \times 10^4 \text{ J}$. (The work is negative because the gas is being compressed.)
- b** From the first law, $Q = \Delta U + W$, so to find the energy transferred we must first find the change in the internal energy of the gas. Since:

$$U = \frac{3}{2}NkT \quad \text{or} \quad U = \frac{3}{2}pV$$

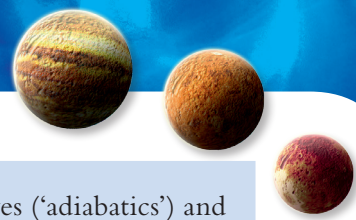
it follows that $\Delta U = \frac{3}{2}(pV)_{\text{final}} - \frac{3}{2}(pV)_{\text{initial}}$. Here the pressure is constant, so this simplifies to:

$$\begin{aligned} \Delta U &= \frac{3}{2}p \Delta V \\ &= \frac{3}{2}(-6.00 \times 10^4) \\ &= -9.00 \times 10^4 \text{ J} \end{aligned}$$

Finally,

$$\begin{aligned} Q &= -9.00 \times 10^4 - 6.00 \times 10^4 \\ &= -1.50 \times 10^5 \text{ J} \end{aligned}$$

The negative sign of Q means that this energy was removed from the gas.



B.20 Figure B.32 is a pressure–volume diagram for an ideal gas, showing two adiabatic curves (‘adiabatics’) and an isovolumetric and an isobaric process, making up the loop ABCD.

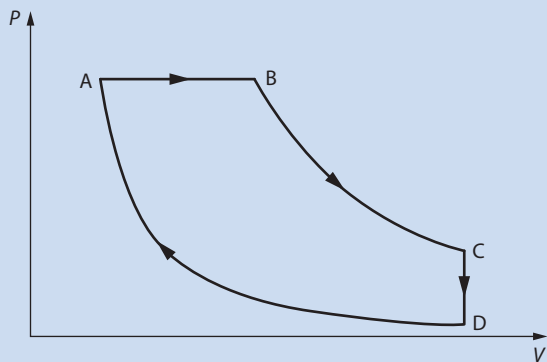


Figure B.32

- Determine along which legs energy is supplied to (Q_{in}) or removed from (Q_{out}) the gas.
- Find the relation between Q_{in} , Q_{out} and the net work done in the loop.

- There is no energy transferred along the adiabatics BC and DA. Along AB, work is done **by** the gas, so $W > 0$, and $\Delta U > 0$ since the temperature at B is higher than that at A. Hence $Q > 0$ and energy is supplied to the gas. Along CD, $W = 0$ (the volume does not change), and $\Delta U < 0$ because the temperature at D is lower than that at C. Hence $Q < 0$ and energy is removed from the gas.
- Applying the first law to the total change $A \rightarrow A$, we see that $\Delta U = 0$, so:

$$Q = 0 + W$$

$$(Q_{\text{in}} - Q_{\text{out}}) = W$$

leading to $W = Q_{\text{in}} - Q_{\text{out}}$. (W is the area of the loop.)

B.21 The loop ABC in Figure B.33 consists of an isobaric, an isovolumetric and an isothermal process for a fixed quantity of an ideal gas.

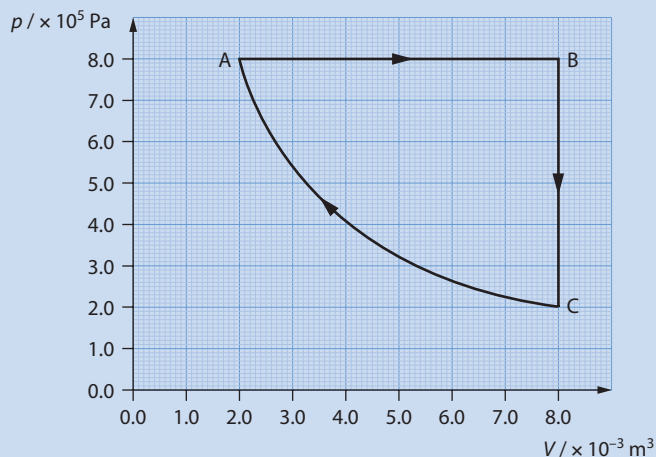


Figure B.33 For Worked example B.21.

The temperature of the gas at A is 320 K. Calculate:

- the temperature at B
- the energy transferred from A to B
- the energy transferred from B to C
- the net work done in one cycle. (The work done on the gas from C to A is 2.2 kJ)

a The temperature at B is found from:

$$\frac{V_A}{T_A} = \frac{V_B}{T_B}$$
$$\frac{2.0 \times 10^{-3}}{320} = \frac{8.0 \times 10^{-3}}{T_B}$$
$$\Rightarrow T_B = 1280 \approx 1300 \text{ K}$$

b The work done by the gas as it expands from A to B is:

$$W_{AB} = P_A \Delta V$$
$$= 8.0 \times 10^5 \times (8.0 \times 10^{-3} - 2.0 \times 10^{-3})$$
$$= 4800 \text{ J} = 4.8 \text{ kJ}$$

From the first law, $Q_{AB} = \Delta U_{AB} + W_{AB}$, so we need to find the change in internal energy. Since the pressure is constant, $\Delta U = \frac{3}{2} P \Delta V = \frac{3}{2} \times 4.8 \times 10^3 = 7.2 \times 10^3 \text{ J}$. Hence $Q_{AB} = 4.8 + 7.2 = +12 \text{ kJ}$.

c $Q_{BC} = \Delta U_{BC}$ since $W = 0$ for the change from B to C (the volume is constant). The magnitude of the temperature change from B to C is the same as that from A to B, so the changes in internal energy have the same magnitude but opposite signs. Therefore $Q_{BC} = -7.2 \text{ kJ}$.

d The net work is therefore $4.8 - 2.2 = 2.6 \text{ kJ}$.



The entropy of the universe is always increasing

The entropy of the universe increases, and we are led from ordered to less ordered and more chaotic situations. But the world around us is full of examples of systems that evolve from highly disordered to very ordered forms (and hence to forms with lower entropy). Life itself evolved from simple microorganisms to the complex forms we see today. A small quantity of water freezing leads to highly structured, ordered (and beautiful) snowflakes. For this to happen, energy has been exchanged with the surroundings in such a way that the systems evolve into highly ordered, complex forms. However, the decrease in entropy in these systems is accompanied by a larger increase in the entropy of the surroundings, leading to an overall increase in the entropy of the universe.

B2.6 The second law of thermodynamics

There are many processes in thermodynamics that are consistent with the first law (energy conservation) but have never been observed to occur. Two examples are:

- the transfer of energy (heat) from a cold body to a hotter body without the performance of work (for example, a glass of water at room temperature freezing, causing the temperature in the rest of the room to rise)
- the air in a room suddenly occupying just one half of the room and leaving the other half empty.

These processes do not happen because they are forbidden by a very special law of physics: the **second law of thermodynamics**. They involve a new concept, that of **entropy**.

Consider an isolated system that is left to change on its own without any intervention from its surroundings. For example, consider a cup of hot coffee left in a cold room. (The system is the cup of coffee and the room.) The natural direction of how things will proceed involves heat leaving the hot coffee and entering the colder room. The coffee will cool down. This is an irreversible process. The reverse will not happen without intervention from the outside. All natural processes are irreversible. An irreversible process captured on film would look absurd if the film were to be run backwards. It looks like thermodynamics is related to the **arrow of time** – the flow of events from the past into the future.

Irreversibility can be quantified. Entropy, like internal energy, is a **state function**: that is, once the state of the system is specified, so is its entropy. Entropy depends only on the state of the system and not on how it got there. What entropy really is and how it is defined is beyond the level of this course. For our purposes, entropy will be taken to be a measure of the disorder of a system.



Time and entropy

The nature of time has always mystified scientists and non-scientists alike. The apparent connection between the direction from past to future (the arrow of time) and thermodynamics is a fascinating and not well-understood aspect of the second law.



The arrow of time

The laws of mechanics do not reveal the arrow of time: a film of billiard balls colliding with each other does not look strange if run backwards. So why do the same laws – when applied to very many particles in a gas, say – show a preference for one direction of time, the one leading to increased entropy

and more disorder? The answer has to do with the fact that the system is led to a more disordered, higher-entropy state because this state is the vastly more likely state, the result of the very many, random collisions among the particles of the gas.

What do we mean by disorder?

- A liquid is more disordered than a solid at the same temperature because its atoms move about, whereas those of the solid are regularly arranged – we know less about the position of the particles in the liquid.
- One mole of a gas in a large volume is more disordered than a mole of the same gas at the same temperature in a smaller volume – we know less about the position of the particles in the larger volume.
- One mole of a gas in a given volume at high temperature is more disordered than one mole in the same volume but at a lower temperature – we know less about the position of the particles at high temperature because they move faster.

In all these examples, increased disorder seems to be associated with lack of information about the system.

Even though we will not define entropy, we can define the change in entropy.

The change ΔS in entropy is defined as:

$$\Delta S = \frac{Q}{T}$$

where Q is a quantity of heat given to or removed from a system at a given temperature T (in kelvin). The unit of entropy is J K^{-1} .

Exam tip

The formula $\Delta S = \frac{Q}{T}$ must be used with great care. Strictly speaking, it applies when a quantity of heat enters or leaves a system without appreciably changing its temperature. If the temperature changes, calculus is required, to evaluate

$$\Delta S = \int \frac{dQ}{T}$$

When heat is given to a system, $Q > 0$ and its entropy increases. When heat is removed, $Q < 0$ and its entropy decreases. For a reversible process that returns the system to its original state, $\Delta S = 0$. One such reversible process is an isothermal expansion of a gas and a subsequent isothermal compression back to the initial state. Since the expansion and compression are isothermal, leaving the temperature constant, we may write:

$$\Delta S_1 = +\frac{|Q|}{T}$$

during expansion and

$$\Delta S_2 = -\frac{|Q|}{T}$$

during compression. (The gas receives thermal energy upon expansion and discards thermal energy upon compression – use the first law of thermodynamics.) The net entropy change is thus zero.

Let us apply the expression for ΔS given above to the case of the flow of heat between a hot body at temperature T_h and a cold body at temperature T_c (Figure **B.34a**). If a quantity of heat Q flows from the hot to the cold body, the total entropy change of the system and its surroundings is:

$$\begin{aligned}\Delta S &= -\frac{Q}{T_h} + \frac{Q}{T_c} \\ &= Q\left(\frac{1}{T_c} - \frac{1}{T_h}\right) \\ &\Rightarrow \Delta S > 0\end{aligned}$$

(the temperature of each body is assumed unchanged during this small exchange of thermal energy). The entropy of the system has increased, so this is what we expect to happen. The reverse process (Figure **B.34b**) does not happen; this corresponds to a net decrease in the entropy of the system and its surroundings:

$$\begin{aligned}\Delta S &= \frac{Q}{T_h} - \frac{Q}{T_c} \\ &= -Q\left(\frac{1}{T_c} - \frac{1}{T_h}\right) \\ &\Rightarrow \Delta S < 0\end{aligned}$$



Entropy and information

Entropy in thermodynamics may be defined as:

$$S = k \ln \Omega$$

where k is Boltzmann's constant and Ω is the number of ways in which a particular macroscopic state of a system may be realised. It is quite extraordinary that this formula of thermodynamics has featured extensively in Claude Shannon's theory of information, a crucial part of modern telecommunications theory.

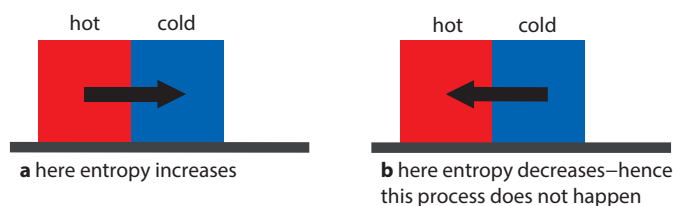
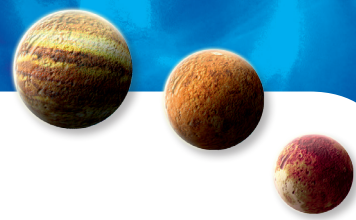


Figure B.34 **a** When thermal energy flows from a hot to a cold body, the entropy of the universe increases. **b** If the reverse were to happen without any performance of work, the entropy would decrease, violating the second law.



Similarly, when energy is given to a solid at its melting temperature, the solid will use that energy to turn into a liquid at the same temperature. The entropy formula again shows that the entropy increases as the solid absorbs the latent heat of fusion.

This allows us to state the second law of thermodynamics in its general form, involving entropy:

The entropy of an isolated system never decreases. In such a system, entropy increases in realistic irreversible processes and stays the same in theoretical, idealised reversible processes.

Worked example

B.22 An ideal gas in state A can reach state B at constant volume or state C at constant pressure (Figure B.35). The temperatures at B and C are the same. Determine which process results in the greater entropy change.

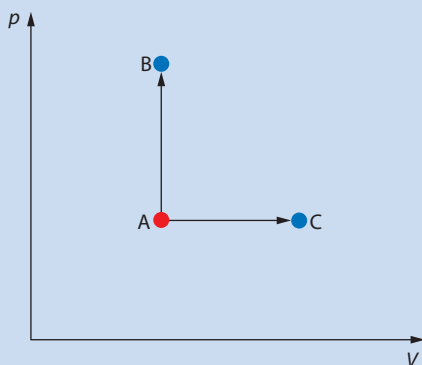


Figure B.35

Since the temperatures at B and C are the same, the change in internal energy is the same for each path. The path AB involves zero work done but the path AC involves positive work done by the gas. Hence Q is greater for the path to C; from $\Delta S = \frac{Q}{T}$, so is the entropy change.

B2.7 Heat engines

It is possible (and easy) to convert mechanical energy into thermal energy: just think of a car when the brakes are applied. The kinetic energy of the car (mechanical energy) is converted into thermal energy in the brake pads: their temperature increases. All of the mechanical energy can be converted into thermal energy in this way. Is the reverse possible? In other words, can we convert thermal energy into mechanical energy with 100% efficiency? Thermodynamics says that this is not possible, as we will see.

A **heat engine** is a device that converts thermal energy into mechanical work. A schematic example is shown in Figure B.36.

A hot reservoir (a source of thermal energy) at temperature T_H transfers heat into the engine. Think of the engine as a gas in a cylinder with a piston. The gas expands, pushing the piston out, and this can be exploited to do mechanical work. Some of the energy transferred into the engine is rejected into a sink at a lower temperature, T_C . The work

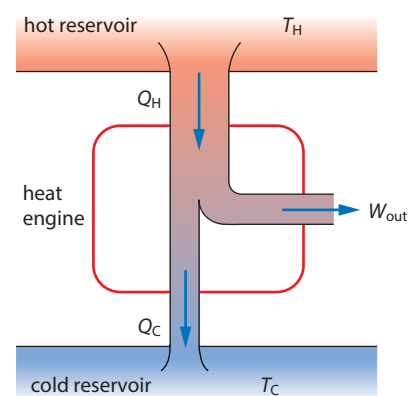


Figure B.36 In a heat engine, energy flowing from the hot to the cold reservoir may be used to perform mechanical work.

done is the heat in, Q_H , minus the heat out, Q_C . Figure B.37 shows a ‘practical’ example of the schematic diagram of Figure B.36.

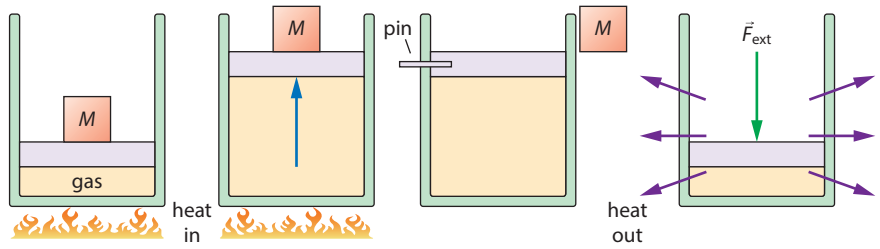


Figure B.37 The hot reservoir provides heat to the gas, which expands, raising a load. The load is removed and the gas is allowed to cool at constant volume. Heat is released to the surroundings. An external force compresses the gas back to its original state. The process can then be repeated, converting heat into mechanical work.

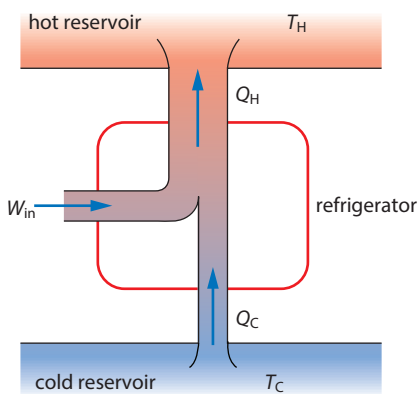


Figure B.38 In a refrigerator, mechanical work is used to extract heat from a cold reservoir and deposit it in the hot reservoir.

Exam tip

Notice that the Kelvin version refers to a heat engine working in a cycle. In an isothermal expansion, **all** the heat supplied to the gas goes into work. But this is not a cyclic process and so does **not** violate the Kelvin version of the second law.

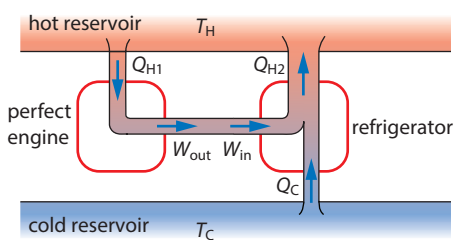


Figure B.39 If a heat engine existed that did not reject heat, then it could be used along with a refrigerator to transfer energy from a cold to a hot reservoir without performing work, violating the second law of thermodynamics (in its Clausius form).

A refrigerator (Figure B.38) is another type of heat engine, in which the performance of mechanical work extracts heat from the cold reservoir and deposits it in the hot reservoir. (The food that is to be kept cool is in the cold reservoir.)

B2.8 Other formulations of the second law

There are a number of formulations of the second law in terms of heat engines, rather than in terms of entropy.

The Clausius version states that:

It is impossible for thermal energy to flow from a cold to a hot object without performing work.

The Clausius version says that you cannot have a ‘workless’ refrigerator: you must do work to push heat from a colder to a hotter object.

The Kelvin version states that:

It is impossible, in a **cyclic process, to completely convert heat into mechanical work.**

The Kelvin version says that you cannot have a perfect engine: any heat engine has to reject some heat into the surroundings and so no heat engine can be 100% efficient.

These are equivalent formulations. If we accept one, we can deduce the other. With more work it can also be shown that these are equivalent to the formulation in terms of entropy. For example, let us suppose that it is possible to have a perfect heat engine – one that does not reject some heat into the surroundings. If such a perfect engine existed, we could couple it to a refrigerator, as shown in Figure B.39. The work done by the perfect engine would be input to the refrigerator. The net effect of the combined system would be the transfer of energy from a cold to a hot reservoir without work. That is impossible, as it would violate the Clausius version of the second law. Hence, no perfect heat engine exists, which is the Kelvin version of the second law.

B2.9 The Carnot cycle

In 1824 the French engineer Sadi Carnot investigated how to convert heat into useful mechanical work in the most efficient way. We know that we cannot have a 100% efficient engine, but is there a limit to how efficient an engine can be? Carnot showed that there is.

Figure B.40 shows the thermodynamic cycle that Carnot investigated. The **Carnot cycle** consists of two isothermals and two adiabatics.

The engine starts its cycle in state 1. The gas is compressed isothermally to state 2. During this stage, heat Q_C leaves the gas and work is done on the gas. From state 2 to state 3 the gas is compressed adiabatically; work is again done on it. From state 3 to state 4 the gas expands isothermally, receiving heat Q_H from a hot reservoir; work is done by the gas. From state 4 to state 1 the gas expands adiabatically; work is done by the gas. The net work done by the engine is therefore:

$$W = Q_H - Q_C$$

The temperature along the isothermal $3 \rightarrow 4$ is T_H , and along the isothermal $1 \rightarrow 2$ it is T_C . The total change in entropy of the engine is:

$$\Delta S = \Delta S_{12} + \Delta S_{23} + \Delta S_{34} + \Delta S_{41}$$

But stages $2 \rightarrow 3$ and $4 \rightarrow 1$ are adiabatics: $Q = 0$, so there is no change in entropy during these stages. Along $3 \rightarrow 4$, $\Delta S_{34} = \frac{Q_H}{T_H}$, and along $1 \rightarrow 2$, $\Delta S_{12} = \frac{Q_C}{T_C}$. Because the process starts and ends at A and because entropy is a state function, the total entropy change is zero: $0 = \frac{Q_H}{T_H} - \frac{Q_C}{T_C}$, which implies that $\frac{Q_H}{T_H} = \frac{Q_C}{T_C}$.

The efficiency of *any* engine is, as usual, $\eta = \frac{\text{useful work}}{\text{input energy}}$. Here the input energy is Q_H , so:

$$\eta_C = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} = 1 - \frac{Q_C}{Q_H}$$

But $\frac{Q_C}{Q_H} = \frac{T_C}{T_H}$ (for the Carnot engine **only**), so we find that $\eta_C = 1 - \frac{T_C}{T_H}$.

Carnot's engine has an efficiency less than 1 (that is, 100%). It would be 100% only in the impossible cases of $T_C = 0$ or $T_H = \infty$.

It can be shown that the following statement is yet another formulation of the second law of thermodynamics:

No engine is more efficient than a Carnot engine operating between the same temperatures.

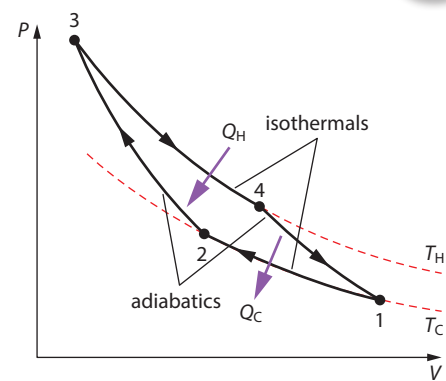


Figure B.40 The Carnot cycle.

Worked example

B.23 Comment on the engine in Figure B.41.

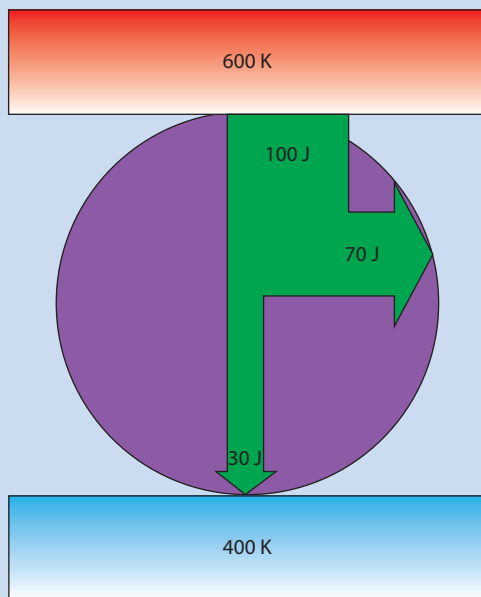


Figure B.41

The efficiency of this engine would be 70%. A Carnot engine operating between the same temperatures would have an efficiency of $1 - \frac{400}{600} = 33\%$, and thus lower, so this one is impossible.

Nature of science

Different views of the same problem

The second law of thermodynamics stands out as a special law of physics. A law such as the conservation of energy places restrictions on what can and cannot happen: a 1.0 kg ball with total energy 10 J cannot jump over a barrier that is higher than 1.0 m; therefore such an event is completely forbidden (in classical physics). The second law of thermodynamics says that certain processes do not happen not because they **really** are impossible, but because they are so highly unlikely as to be impossible for all practical purposes. Different scientists working on the relationship between heat and work framed this law in different ways, as they were working from different perspectives. Carnot drew his conclusions about the limitations on the efficiencies of heat engines before the concept of entropy was discovered. Through testing and collaboration, scientists realised that the alternative statements were equivalent. The connection of the second law with probability and the number of microstates that realise a particular macroscopic state makes this law a truly special law of physics.

? Test yourself

- 18 An ideal monatomic gas expands adiabatically from a state with pressure $8.1 \times 10^5 \text{ Pa}$ and volume $2.5 \times 10^{-3} \text{ m}^3$ to a state of volume $4.6 \times 10^{-3} \text{ m}^3$. Calculate the new pressure of the gas.
- 19 An ideal monatomic gas expands adiabatically from a state with volume $2.8 \times 10^{-3} \text{ m}^3$ and temperature 560 K to a state of volume $4.8 \times 10^{-3} \text{ m}^3$. Calculate the new temperature of the gas.
- 20 A gas expands at a constant pressure of $5.4 \times 10^5 \text{ Pa}$ from a volume of $3.6 \times 10^{-3} \text{ m}^3$ to a volume of $4.3 \times 10^{-3} \text{ m}^3$. Calculate the work done by the gas.
- 21 A gas is compressed isothermally so that an amount of work equal to 6500 J is done on it.
- Calculate how much energy is removed from or given to the gas.
 - The same gas is instead compressed adiabatically to the same final volume as in **a**. Suggest whether the work done on the gas will be less than, equal to or greater than 6500 J . Explain your answer.
- 22 An ideal gas expands isothermally from pressure P and volume V to a volume $2V$. Sketch this change on a pressure–volume diagram. An equal quantity of an ideal gas at pressure P and volume V expands adiabatically to a volume $2V$. Sketch this change on the same axes. Determine in which case the work done by the gas is greater.
- 23 An ideal gas is compressed isothermally from pressure P and volume $2V$ to a volume V . Sketch this change on a pressure–volume diagram. An equal quantity of an ideal gas at pressure P and volume $2V$ is compressed adiabatically to a volume V . Sketch this change on the same axes. Determine in which case the work done on the gas is greater.

- 24 A quantity of energy Q is supplied to three ideal gases, X, Y and Z. Gas X absorbs Q isothermally, gas Y isovolumetrically and gas Z isobarically. Copy the table below and complete it by inserting the words ‘positive’, ‘zero’ or ‘negative’ for the work done W , the change in internal energy ΔU and the temperature change ΔT for each gas.

W	ΔU	ΔT
X		
Y		
Z		

- 25 An ideal gas is compressed adiabatically.
- Use the first law of thermodynamics to state and explain the change, if any, in the temperature of the gas.
 - Explain your answer to **a** by using the kinetic theory of gases.
- 26 An ideal gas is kept at constant pressure of $6.00 \times 10^6 \text{ Pa}$. Its initial temperature is 300 K . The gas expands at constant pressure from a volume of 0.200 m^3 to a volume of 0.600 m^3 . Calculate:
- the work done by the gas
 - the temperature of the gas at the new volume
 - the energy taken out of or put into the gas.
- 27 An amount Q of energy is supplied to a system. Explain how it is possible that this addition of thermal energy might **not** result in an increase in the internal energy of the system.
- 28 An ideal gas undergoes a change from state P to state Q, as shown below (the temperature is in kelvin). For this change, state and explain:
- whether work is done on the gas or by the gas
 - whether energy is supplied to the gas or taken out of the gas.

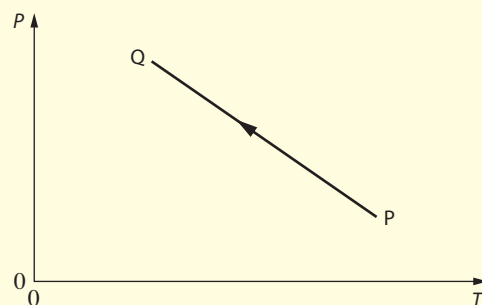
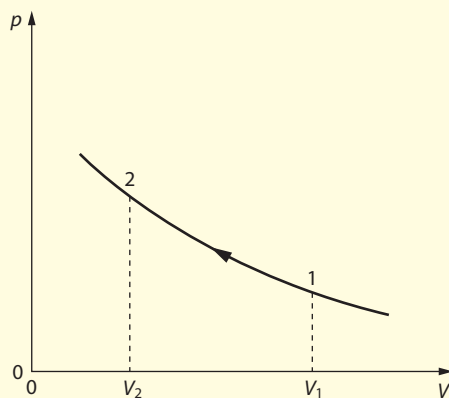


Figure B.52

29 Two ideal gases, X and Y, have the same pressure, volume and temperature. The same quantity of energy is supplied to each gas. Gas X absorbs the thermal energy at constant volume, whereas gas Y absorbs the thermal energy at constant pressure. State and explain which of the two gases will have the larger final temperature.

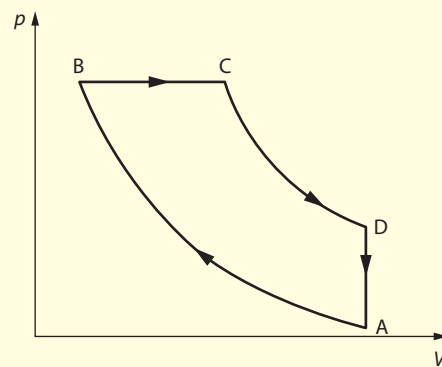
30 Two ideal gases, X and Y, have the same pressure, volume and temperature. A quantity of energy is supplied to each gas. Gas X absorbs the thermal energy at constant volume, whereas gas Y absorbs the thermal energy at constant pressure. The increase in temperature of both gases is the same. State and explain which of the two gases received the larger quantity of energy.

31 The graph below shows the isothermal compression of an ideal gas from state 1 of volume V_1 to state 2 of volume V_2 .



- Copy the figure and on it draw a curve to show the adiabatic compression of the same gas from state 1 to a different state of volume V_2 .
- Explain in which case the work done on the gas is the larger.
- The temperature of the gas in state 1 is 300 K. The temperature of the surroundings may be assumed constant at 300 K. The work done on the gas during the isothermal compression is 25 kJ. Determine the change in entropy:
 - of the gas during the isothermal compression
 - of the surroundings during the isothermal compression.
- Discuss how the answers to c are consistent with the second law of thermodynamics.

32 The graph below shows a cycle of a heat engine working with an ideal gas. Curves AB and CD are adiabatics.



- State what is meant by an adiabatic curve.
 - State the names of the processes BC and DA.
 - Determine along which stage heat is given to the gas.
 - Along DA the pressure drops from 4.0×10^6 Pa to 1.4×10^6 Pa. The volume along DA is 8.6×10^{-3} m³. The efficiency of the engine is 0.36. The number of moles of gas is 0.25. Calculate:
 - the heat taken out of the gas
 - the heat given to the gas
 - the area of the loop ABCD.
- 33 The **molar specific heat capacity** of an ideal gas is defined as the amount of energy required to change the temperature of one mole of the gas by 1 K. When n moles of gas absorb energy Q at constant pressure,

$$Q = nc_p \Delta T$$

where c_p is the molar specific heat capacity at constant pressure.

When n moles of gas absorb energy Q at constant volume,

$$Q = nc_v \Delta T$$

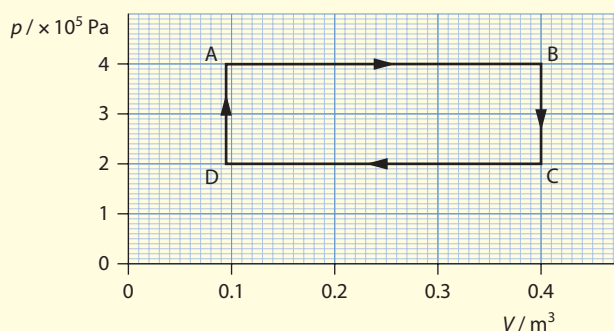
where c_v is the molar specific heat capacity at constant volume.

Use the first law of thermodynamics to show that:

$$c_p - c_v = R$$

where R is the universal gas constant.

- 34 Consider the loop ABCD as shown on the graph. The temperature at A is 800 K.



Calculate:

- the temperature at points B, C and D
 - the change in internal energy along each of the four legs of the cycle
 - the amount of energy given to and taken from the gas along each leg
 - the efficiency of the cycle.
- 35 A quantity of ice at -10°C is dropped into water. All the ice eventually melts and the final temperature of the water is $+15^\circ\text{C}$. Describe and compare the entropy changes taking place in the ice and in the water during the following processes. (Assume that there are no energy exchanges between the ice–water system and the surroundings.)
- The temperature of the ice increases from -10°C to 0°C .
 - The ice is melting.
 - All the ice has melted and the water is approaching its final temperature of $+15^\circ\text{C}$.
- 36 Explain whether you should invest money in a revolutionary new heat engine whose inventor claims that it operates between temperatures of 300 K and 500 K with an efficiency of 0.42.
- 37 **a** State the Clausius form of second law of thermodynamics.
b Show that if a heat engine had a greater efficiency than the Carnot efficiency the second law in a would be violated. (Imagine the work output of the engine to be the input work in a Carnot refrigerator.)

B3 Fluids (HL)

This section deals with fluids at rest and fluids in motion. The equilibrium and motion of fluids are of importance in very many different disciplines. Understanding how fluids behave is crucial in the design of aircraft and cars, in monitoring blood flow in the arteries of a patient, in the efficient use of water by agricultural engineers – but also to a theoretical astrophysicist working with models of stars. We will meet Archimedes' important principle about buoyant forces and a few of the many applications of the Bernoulli equation.

B3.1 Pressure

Consider a fluid in a container (Figure B.42). The surface of the fluid is exposed to the atmosphere. The pressure of the fluid at the top surface is equal to atmospheric pressure. As we move deeper into the fluid, the pressure increases, because of the weight of the fluid above each given point. Figure B.42 shows a region of the fluid marked by a dashed line. The fluid inside the line is in equilibrium, so the net force on it is zero. The forces on this region are the weight of the fluid within it, the force due to the pressure at the top surface and the pressure at the lower surface.

We denote atmospheric pressure by $p_{\text{atm}} = p_0$. Atmospheric pressure varies depending on location; the standard value at sea level is taken to be $1.0 \times 10^5 \text{ Pa}$. Remember that $p = \frac{F}{A} \Rightarrow F = pA$. Equilibrium requires that: (A is the area of the top and lower surfaces of the fluid and mg the weight of the fluid in the marked region)

$$pA = p_0A + mg$$

Learning objectives

- Work with density and pressure.
- Apply Archimedes' principle.
- Apply Pascal's principle.
- Apply hydrostatic equilibrium.
- Appreciate the ideal fluid.
- Understand streamlines.
- Apply the continuity equation and the Bernoulli equation.
- Apply Stokes' law and understand viscosity.
- Distinguish between laminar and turbulent flow.
- Appreciate the meaning of the Reynolds number.

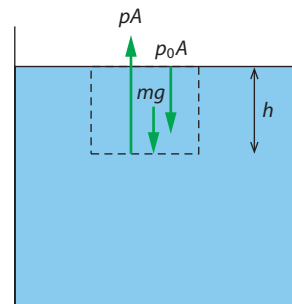


Figure B.42 The pressure increases as the depth h increases.

But:

$$\begin{aligned} mg &= \rho Vg \\ &= \rho Ahg \end{aligned}$$

where ρ is the density of the liquid. Hence:

$$\begin{aligned} pA &= p_0A + \rho Ahg \\ p &= p_0 + \rho gh \end{aligned}$$

We see that, at a depth h below the free surface of the fluid, the pressure depends both on the quantity ρgh , associated with the weight of the fluid in the dotted region, and on the atmospheric pressure at the top surface.

It is important to note that the pressure at the same depth of a connected fluid in equilibrium is the same no matter what the shape of the container (Figure B.43).

This is the basis for an instrument, called a **manometer**, for measuring pressure (Figure B.44).

The U-shaped tube is filled with a liquid of density ρ . The difference in the liquid levels in the two columns is h . The pressure of the liquid is equal anywhere along the dashed horizontal line. On the left side, the pressure is equal to the pressure p of the gas. On the right side, it is equal to $p_0 + \rho gh$, where p_0 is atmospheric pressure. These are equal, so $p = p_0 + \rho gh$.

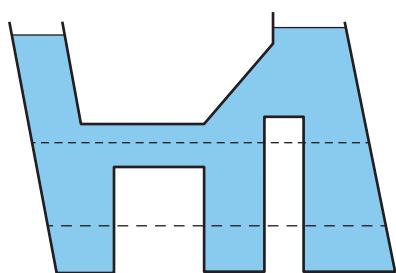


Figure B.43 The pressure is the same at any point along a given horizontal dashed line.

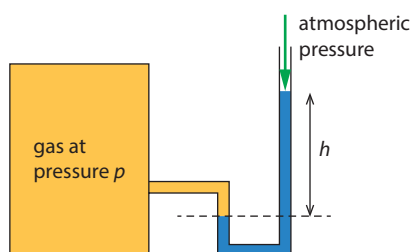


Figure B.44 Using a manometer, the pressure of a gas may be determined from the value of atmospheric pressure and the difference in height of the two columns.

B3.2 Pascal's principle

Pascal's principle states that, when pressure is applied to any point of an **enclosed, incompressible** fluid, the pressure will be transmitted to all other parts of the liquid and the walls of the container of the fluid.

Figure B.45 shows an enclosed fluid. A force F_1 is applied to a piston of area A_1 . The pressure in the liquid immediately below the piston is therefore $p_1 = \frac{F_1}{A_1}$. This is also the pressure right under the second, larger piston, because they are at the same horizontal level. But the pressure there is also $p_2 = \frac{F_2}{A_2}$. Hence:

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}$$

This shows that a small force F_1 applied to a small piston can give rise to a larger force on a larger piston – lifting a heavy object such as a car in the case of a **hydraulic lift**. Note, however, that the smaller force must push through a larger distance, d_1 in order to lift the heavy load by a distance d_2 such that $F_1 \times d_1 = F_2 \times d_2$.

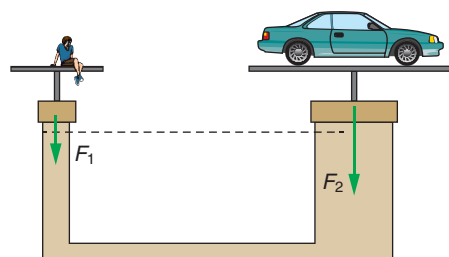


Figure B.45 A hydraulic lift: a small force gets multiplied and lifts a heavy object.

B3.3 Archimedes' principle

A body that is totally or partially immersed in a fluid will experience an upward force called the buoyant force. Consider a region of a fluid that has been marked. The fluid is in equilibrium, so its weight is balanced by the buoyant force (Figure B.46). The buoyant force is the result of the pressure acting on the body from all directions.

Because the pressure is higher at greater depths, the net force exerted by the surrounding fluid is directed upwards, and is just equal to the weight of the fluid in the marked region (otherwise the fluid would move). Now replace the marked fluid with a body that just fits that space. The buoyant force remains equal to the weight of the displaced fluid. The arrows representing the pressure are the same as before, so the net force they give rise to is the same as before; it equals the weight of the displaced fluid. This is **Archimedes' principle**:

A body partially or completely immersed in a fluid experiences a buoyant force B that equals the weight of the displaced fluid.

The weight of the displaced fluid is mg , where $m = \rho V_{\text{displ}}$, so

$$B = \rho g V_{\text{displ}}$$

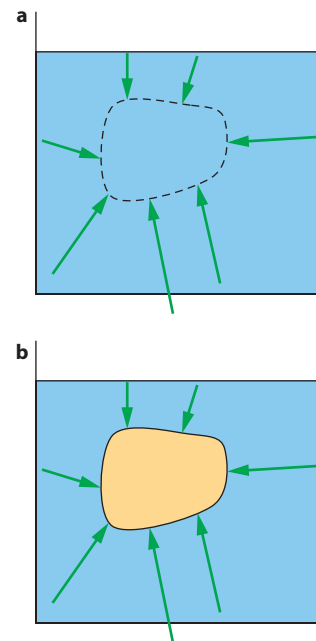


Figure B.46 **a** Fluid pressure acts on all sides of the marked region of the liquid. It is larger at greater depths, so the net force from the surrounding fluid is upwards, just balancing the weight of the fluid in the marked region. **b** If the fluid in the marked region is replaced by another body, the buoyant force remains unchanged, and equal to the weight of the displaced fluid.

Worked example

B.24 Figure B.47 shows a floating rectangular platform made out of wood of density 640 kg m^{-3} . The fluid is water of density 1000 kg m^{-3} .

Calculate the fraction $\frac{h}{d}$.

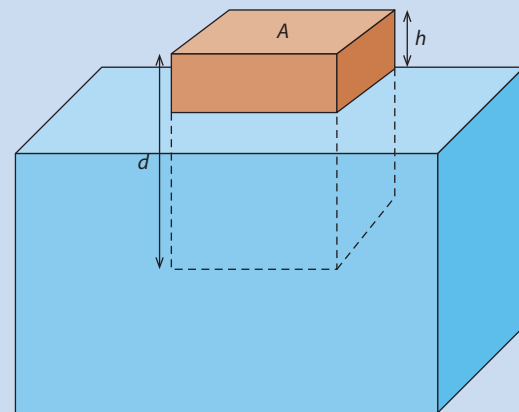


Figure B.47

Let A denote the surface area of the wood. We have equilibrium, so weight equals buoyant force:

$$W = B$$

$$\rho_{\text{wood}} A d g = \rho_{\text{water}} A (d - h) g$$

$$640 d = 1000 (d - h)$$

$$1000 h = 360 d$$

$$\frac{h}{d} = 0.36$$

B3.4 The ideal fluid

In previous subsections, we considered fluids at rest. We will now examine fluids in motion. The motion of fluids is very complex. To simplify the analysis, a number of assumptions about fluids are made. These assumptions define an **ideal fluid**. This is a fluid whose flow is:

- **Steady:** At any fixed point in space, the fluid velocity does not change with time. Steady flow is also called **laminar flow**.
- **Incompressible:** The density is the same everywhere in the fluid.
- **Non-viscous:** A body moving through the fluid would feel no resistive drag forces.

Consider an ideal fluid that moves. We may concentrate on a very small bit of the fluid – a fluid element – and follow its motion. This fluid element traces out a path as it moves (Figure B.48). We call this path a **streamline**. The velocity of the fluid element is tangent to the streamline. Streamlines cannot cross. A set of neighbouring streamlines forms a **flowtube**. A flowtube has streamlines as its boundary.

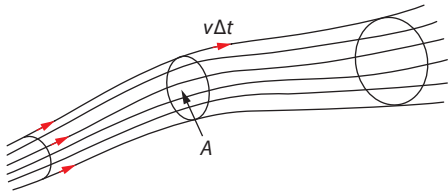


Figure B.48 Streamlines and a flowtube. The velocity vectors are tangent to streamlines.

B3.5 The continuity equation

We know that the speed of water out of a garden hose increases if we use a finger to partially close the hose opening. This is a consequence of the **equation of continuity**. Suppose we have a fluid that flows in a tube of uniform cross-sectional area A (Figure B.49a). The volume of fluid that can go through the dashed line in time Δt is $A v \Delta t$.

This volume will enter the region shaded darker blue. Since the fluid is incompressible, an equal volume of fluid must leave the blue shaded area during the same time in order to keep the **density** constant. If the fluid flows in a tube of varying cross-sectional area (Figure B.49b), we must have that:

$$A_1 v_1 \Delta t = A_2 v_2 \Delta t \Rightarrow A_1 v_1 = A_2 v_2$$

So if the cross-sectional area decreases, the fluid speed must increase.

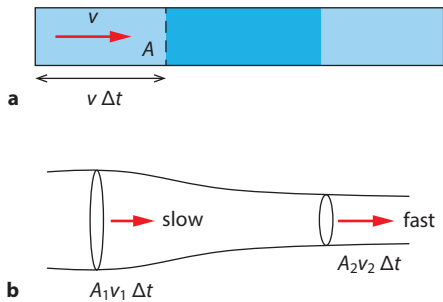


Figure B.49 a The volume of fluid that enters a pipe in time Δt must also leave in the same time from the other end. b This implies that if the cross-sectional area changes the speed must also change.

Worked example

B.25 Water comes out of a tap of cross-sectional area 1.5 cm^2 . After falling a vertical distance of 6.0 cm , the cross-sectional area of the water column has been reduced to 0.45 cm^2 . Calculate the speed of the water as it left the tap.

Let the required speed be v_1 . By the equation of continuity,

$$A_1 v_1 = A_2 v_2$$

where v_2 is the speed after falling 6.0 cm . The water is falling freely under gravity and so:

$$v_2^2 = v_1^2 + 2gh$$

Thus, $A_1 v_1 = A_2 \sqrt{v_1^2 + 2gh}$. Squaring, this gives:

$$A_1^2 v_1^2 = A_2^2 (v_1^2 + 2gh)$$

$$v_1^2 (A_1^2 - A_2^2) = 2A_2^2 gh$$

$$v_1 = \sqrt{\frac{2A_2^2 gh}{A_1^2 - A_2^2}} = \sqrt{\frac{2 \times 0.45^2 \times 9.8 \times 0.060}{1.5^2 - 0.45^2}} = 0.34 \text{ m s}^{-1}$$

B3.6 The Bernoulli equation

The **Bernoulli equation** applies to the laminar flow of a fluid in a tube of varying cross-sectional area and varying vertical height, such as the tube shown in Figure B.50.

This equation relates pressure, height and speed along a streamline. It states that:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g z_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g z_2$$

As we will see, this is a consequence of energy conservation.

In particular, if the height stays constant ($z_1 = z_2$) the equation states that:

$$p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2$$

This shows that in those areas in the flow where the speed is high the pressure is low, and vice versa.

To derive the Bernoulli formula, we start by noticing that the work done by the net force is the change in the kinetic energy of the fluid:

$$\begin{aligned} W_{\text{net}} &= \Delta E_K \\ &= \frac{1}{2}m(v_2^2 - v_1^2) \\ &= \frac{1}{2}\rho \Delta V(v_2^2 - v_1^2) \end{aligned}$$

The net work on the fluid consists of the work done against gravity (its weight) and the work associated with the pressure of the fluid. The work done against gravity is

$$\begin{aligned} W_{\text{weight}} &= -(mgz_2 - mgz_1) \\ &= -\rho g \Delta V(z_2 - z_1) \end{aligned}$$

To calculate the work associated with the pressure of the fluid, recall that work is the product of force times the distance over which it acts. The force on the left end of the volume of fluid in Figure B.50 is $F_1 = p_1 A_1$, where A_1 is the cross-sectional area of the tube there, and the distance the fluid moves is Δx_1 , so the work done at the left end is $F_1 \Delta x_1 = p_1 A_1 \Delta x_1 = p_1 \Delta V_1$. Likewise, at the right end the fluid does work equal to $p_2 \Delta V_2$, or we may say it has $-p_2 \Delta V_2$ of work done on it. But the amount of fluid that enters this section of the tube is the same amount that leaves, so $\Delta V_1 = \Delta V_2$. The net work associated with the pressure of the fluid is thus:

$$W_{\text{pressure}} = p_1 \Delta V - p_2 \Delta V$$

Equating the total work done and the change in kinetic energy gives:

$$p_1 \Delta V - p_2 \Delta V - \rho g \Delta V(z_2 - z_1) = \frac{1}{2}\rho \Delta V(v_2^2 - v_1^2)$$

Rearranging and cancelling ΔV gives the Bernoulli formula. A direct application of the Bernoulli equation is the problem of the flow of a liquid out of a container. Figure B.51 shows a tank of liquid where a hole has been made in the side at a depth h below the liquid surface.

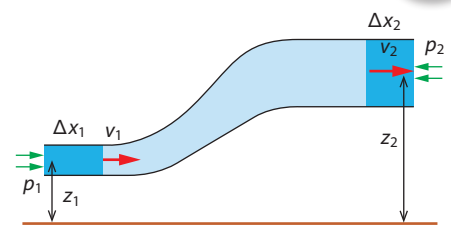


Figure B.50 Diagram for demonstrating the Bernoulli equation.

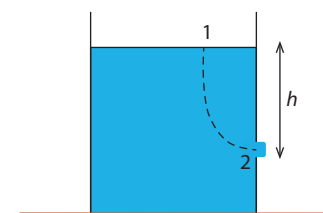


Figure B.51 Fluid flowing out of a container. Points 1 and 2 are connected by a streamline, along which the Bernoulli equation may be applied.

Notice that both the surface and the hole are exposed to the atmosphere and so the pressure there is atmospheric pressure, p_0 . The dashed line shows a possible streamline joining the surface and the hole. Applying the Bernoulli equation to the streamline joining 1 to 2, we find:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g z_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g z_2$$

We measure heights from the level of the hole, so $z_1 = h$ and $z_2 = 0$. The surface has negligibly small speed (either because the surface is very large or because the liquid is being replenished to keep h constant), so the Bernoulli equation simplifies to:

$$p_0 + 0 + \rho g h = p_0 + \frac{1}{2}\rho v_2^2 + 0$$

The equation then gives:

$$v_2 = \sqrt{2gh}$$

for the speed at which the liquid leaves the hole.

Worked examples

B.26 Water of density 1000 kg m^{-3} flows in a horizontal pipe (Figure B.52). The radius of the pipe at its left end is 65 mm and that at the right end is 45 mm. The water enters from the left end with a speed of 6.0 m s^{-1} . The pressure at the left end is 185 kPa. Calculate the pressure at the right end of the pipe, at a vertical distance of 1.5 m above the left end.

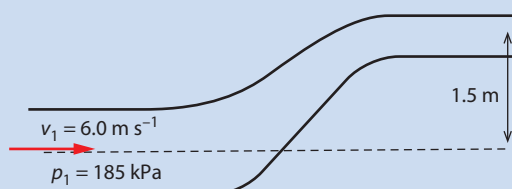


Figure B.52

We apply the Bernoulli equation: $p_1 + \frac{1}{2}\rho v_1^2 + \rho g z_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g z_2$. We have two unknowns, the speed and pressure of the water at the right end. We use the equation of continuity to find the second speed:

$$A_1 v_1 = A_2 v_2 \Rightarrow v_2 = \frac{A_1 v_1}{A_2} = \frac{\pi(65^2) \times 6.0}{\pi(45^2)} = 12.5 \text{ m s}^{-1}$$

Going back to the Bernoulli equation, we now have

$$185 \times 10^3 + \frac{1}{2} \times 10^3 \times 6.0^2 + 0 = p_2 + \frac{1}{2} \times 10^3 \times 12.5^2 + 10^3 \times 9.8 \times 1.5$$

This gives 110 kPa for the pressure p_2 .



B.27 In a hydroelectric power plant (Figure B.53), water leaves a dam from a point 50 m beneath the surface. It enters a pipe of radius 80 cm and is incident on a turbine through a pipe of radius 40 cm. Calculate **a** the speed of the water as it hits the turbine, and **b** the pressure at point 2.

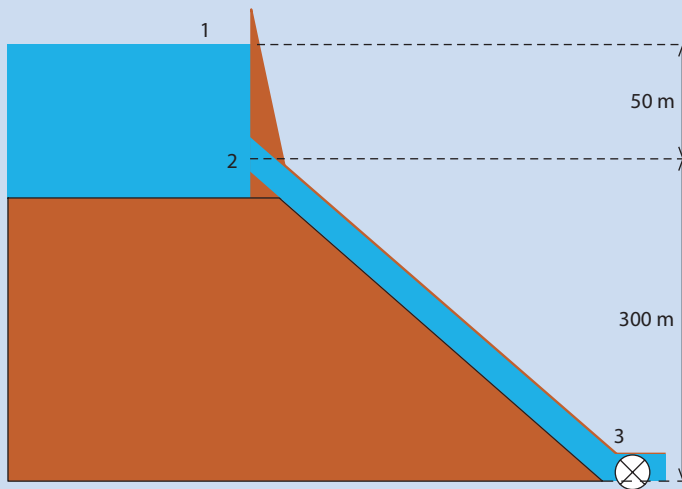


Figure B.53

a We consider a streamline beginning at 1 and ending at 3. The speed at 1 may be taken to be zero; the surface area of the water is assumed large, so the surface does not move appreciably. At 1 and 3 the pressure is atmospheric, since the water is exposed to the atmosphere. Therefore, from the Bernoulli equation,

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g z_1 = p_3 + \frac{1}{2}\rho v_3^2 + \rho g z_3$$

$$p_{\text{atm}} + 0 + 10^3 \times 9.8 \times 350 = p_{\text{atm}} + \frac{1}{2} \times 10^3 \times v_3^2 + 0$$

$$v_3 = 83 \text{ m s}^{-1}$$

b Consider a streamline beginning at 1 and ending at 2. The speed of the water at 2 can be found from the continuity equation relating points 2 and 3:

$$A_2 v_2 = A_3 v_3 \Rightarrow v_2 = \frac{A_3 v_3}{A_2} = \frac{\pi(40^2) \times 83}{\pi(80^2)} = 21 \text{ m s}^{-1}$$

From the Bernoulli equation again,

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g z_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g z_2$$

$$p_{\text{atm}} + 0 + 10^3 \times 9.8 \times 50 = p_2 + \frac{1}{2} \times 10^3 \times 21^2 + 0$$

$$p_2 = 3.7 \times 10^5 \text{ Pa}$$

B3.7 The Bernoulli effect

The classic application of the Bernoulli equation is to the motion of air past an aircraft wing. A wing shape (with a cross-section illustrated in Figure B.54) is called an **aerofoil**. The figure shows streamlines as they move below and above the aerofoil.

Because of the shape of the wing, air flows faster above the wing than beneath it. The streamlines above the wing are closer together, indicating that the air there is moving faster. A flow tube above the aerofoil would have a smaller area, and by the equation of continuity the air must move faster. The larger speed above the aerofoil results in a

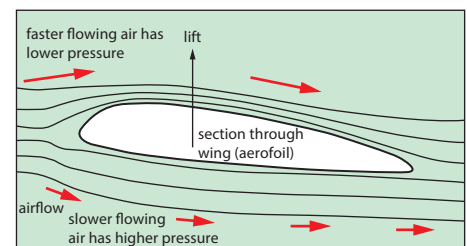


Figure B.54 Fluid flows faster over the top of the aerofoil than beneath it, giving rise to a lifting force.

smaller pressure, so there is a net upward force on the aerofoil, a lifting force. More precisely, from:

$$p_{\text{over}} + \frac{1}{2}\rho v_{\text{over}}^2 + \rho g z = p_{\text{under}} + \frac{1}{2}\rho v_{\text{under}}^2 + \rho g z$$

we have

$$p_{\text{under}} - p_{\text{over}} = \frac{1}{2}\rho v_{\text{over}}^2 - \frac{1}{2}\rho v_{\text{under}}^2 > 0$$

where we have neglected the $\rho g z$ terms, which are almost the same for a thin aerofoil. The difference in pressure results in a lifting force equal to:

$$F = (p_{\text{under}} - p_{\text{over}})A = \frac{1}{2}\rho(v_{\text{over}}^2 - v_{\text{under}}^2)A$$

where A is the area of the aerofoil.

A similar effect occurs for air flowing past the sail of a sailboat. The sail plays the role of an aerofoil, and the difference in speeds of the air on the two sides of the sail gives a force on the sail that propels the sailboat.

The **Bernoulli effect** is often exploited in sports. Figure B.55 shows a ball that has been set spinning as it was thrown. (The ball travels from left to right so air is shown flowing from right to left.) A layer of air is ‘dragged along’ with the spinning surface of the ball. This, added to the overall air flow from the ball’s translational motion, gives a total air speed that is higher on the underside and lower on the top side. The result is a net downward pressure on the ball, giving it an unexpectedly curved path.

Figure B.56 shows a device known as a **Pitot–Prandtl tube**, which can be used to measure the speed of flow of air past an object such as an aircraft.

This is a thin tube with a small opening at the front, pointing in the direction of the aircraft’s motion. Air enters at the front hole at B and, essentially immediately, is brought to rest. Using the Bernoulli equation along streamline BA we have

$$p_A + \frac{1}{2}\rho_{\text{air}}v_A^2 = p_B + \frac{1}{2}\rho_{\text{air}}v_B^2$$

But $v_B = 0$, so:

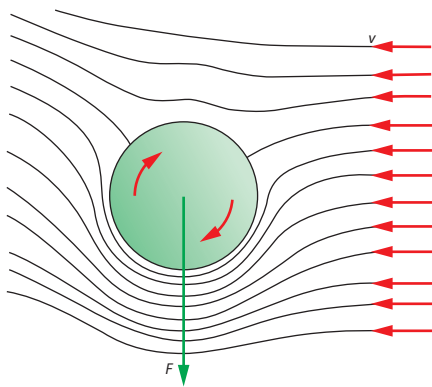
$$p_A + \frac{1}{2}\rho_{\text{air}}v_A^2 = p_B + 0$$

We have neglected the $\rho g z$ terms, which are almost the same for A and B. Rearranging the equation,

$$v_A = \sqrt{\frac{2(p_B - p_A)}{\rho_{\text{air}}}}$$

One leg of a liquid manometer is connected to the front hole of the tube and the other leg to A (there are tiny holes on the sides of the tube). The difference in pressure is measured by the height h between the columns of the manometer: $p_B = p_A + \rho_{\text{liquid}}gh$. Hence:

$$v = \sqrt{\frac{2\rho_{\text{liquid}}gh}{\rho_{\text{air}}}}$$



B.55 The ball spins clockwise as it moves to the right. This means that the air flows faster beneath it, so the ball experiences a downward force F .

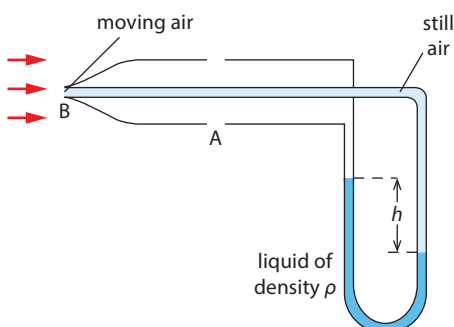
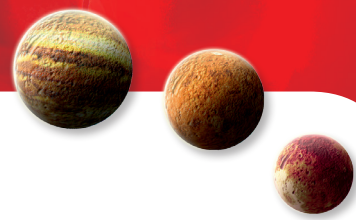


Figure B.56 A Pitot–Prandtl tube used to measure the flow of a fluid past an object such as an aircraft.



Yet another application of the Bernoulli effect is the **Venturi tube** (Figure B.57), which is also used to measure fluid flow speeds.

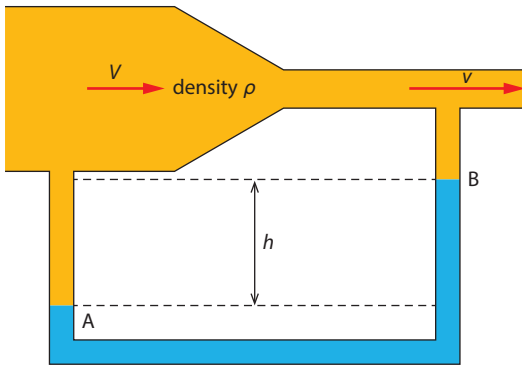


Figure B.57 A Venturi tube, which can be used to measure the flow of a fluid in a pipe.

The constriction in the tube makes the fluid move faster there and so the pressure drops. The wide and narrow parts of the tube are connected to the two arms of a liquid manometer. Because of the difference in pressure the two columns have a difference in height: $p_A = p_B + \rho_{\text{liquid}}gh$, so $\Delta p = p_A - p_B = \rho_{\text{liquid}}gh$. From the Bernoulli equation,

$$p_A + \frac{1}{2}\rho V^2 = p_B + \frac{1}{2}\rho v^2$$

and, from the equation of continuity, $AV = av$, where A and a are the cross-sectional areas of the wide and narrow parts of the tube.

Hence, $v = \frac{AV}{a}$. Substituting this into the Bernoulli equation,

$$p_A + \frac{1}{2}\rho V^2 = p_B + \frac{1}{2}\rho \frac{A^2 V^2}{a^2}$$

and solving for V , the entry speed in the wide tube, we find:

$$p_A - p_B = \frac{1}{2}\rho \left(\frac{A^2}{a^2} - 1 \right) V^2$$

$$V^2 = \frac{2(p_A - p_B)}{\rho \left(\frac{A^2}{a^2} - 1 \right)}$$

$$V^2 = \frac{2a^2(p_A - p_B)}{\rho(A^2 - a^2)}$$

$$V = \sqrt{\frac{2a^2(p_A - p_B)}{\rho(A^2 - a^2)}}$$

$$= \sqrt{\frac{2a^2 \Delta p}{\rho(A^2 - a^2)}}$$

Exam tip

There is a lot of algebra in this derivation and you must study it carefully. In an exam you will most likely be guided in such a derivation.

Notice that most fluid dynamics problems involve the use of both the Bernoulli and the continuity equations.

In these equations ρ is the density of the fluid that flows and ρ_{liquid} is the density of the liquid in the manometer.

Worked example

B.28 In a particular Venturi tube, the wide and narrow parts of the tube have cross-sectional areas of 45 cm^2 and 25 cm^2 , respectively. The pressure difference between the wide and narrow parts of the tube is 65 kPa . Determine the speed of water flowing in the wide part of the tube and the volume flow rate in the tube.

This is a straightforward use of the formula just derived:

$$V = \sqrt{\frac{2a^2 \Delta p}{\rho(A^2 - a^2)}} = \sqrt{\frac{2 \times 25^2 \times 65 \times 10^3}{10^3 \times (45^2 - 25^2)}}$$

$$= 7.6 \text{ m s}^{-1}$$

The volume flow rate is $Q = VA = 7.6 \times 45 \times 10^{-4} = 3.4 \text{ m}^3 \text{ s}^{-1}$

Exam tip

A ball at rest in a fluid undergoes constant collisions with the molecules of the fluid. If the ball is falling in the fluid, the lower side of the ball is moving into the molecules and the top side is moving away from them. This means that the lower side will have more collisions per second and hence will experience a greater force upwards. The greater the speed, the greater the number of collisions; hence it is reasonable that drag force is proportional to speed.

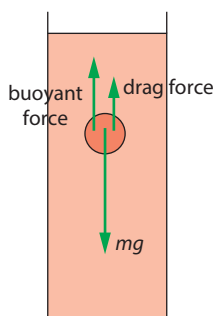


Figure B.58 A sphere falling through a fluid experiences a drag force.

B3.8 Stokes' law and viscosity

Stokes' law states that a sphere falling through an ideal fluid would experience no resistive force (other than the buoyant force). Similarly, the propeller of a ship turning in an ideal fluid would exert no force on the fluid, and consequently the reaction force on the propeller would be zero. In a real fluid, there are of course forces in both cases.

One of the differences between ideal and real fluids is that real fluids have **viscosity**. Viscosity has to do with how a layer of the fluid affects the motion of neighbouring layers. Viscosity is measured in terms of a coefficient with units of Pa s.

A small sphere of radius r falling through a viscous fluid (Figure B.58) experiences a resistive, or drag, force. For small velocities, the drag force is given by $F = 6\pi\eta rv$, where v is the speed and η is the viscosity coefficient.

The net force on a small sphere released from rest in a viscous fluid is:

$$F_{\text{net}} = mg - 6\pi\eta rv - \rho_{\text{fluid}}gV$$

The final term is the buoyant force on the sphere. The sphere will accelerate downwards, but as the speed increases the drag force increases, and at a high enough speed the net force will become zero, and the sphere will continue to move at constant speed. This speed is called the sphere's **terminal speed**. Using $m = \rho_{\text{body}}V$ and $V = \frac{4\pi r^3}{3}$, this speed is given by:

$$mg - 6\pi\eta rv - \rho_{\text{fluid}}gV = 0$$

$$v = \frac{\rho_{\text{body}}Vg - \rho_{\text{fluid}}gV}{6\pi\eta r}$$

$$= \frac{(\rho_{\text{body}} - \rho_{\text{fluid}})2r^2g}{9\eta}$$

B3.9 Turbulence

The top diagram in Figure B.59 shows steady laminar flow: the speed of elements of the fluid along a cross-section of the pipe is constant.

The middle diagram shows viscous laminar flow: the elements of the fluid near the surface of the pipe move more slowly. The lower diagram shows **turbulent flow**: the streamlines are unpredictable and chaotic.

Turbulent flow arises as fluid speed increases.

The transition from laminar to turbulent flow is hard to measure, but a rough estimate is provided by a dimensionless number known as the **Reynolds number**. For a fluid flowing with speed v in a pipe of radius r , this is defined as:

$$R = \frac{v\rho r}{\eta}$$

where ρ is the density of the fluid and η its viscosity. We have turbulent flow if this number exceeds about 1000.

As an example, consider air of density 1.2 kg m^{-3} that flows with speed 2.1 m s^{-1} in a pipe of radius 5.0 mm . The viscosity of the air is $1.8 \times 10^{-5} \text{ Pa s}$. The Reynolds number is:

$$R = \frac{v\rho r}{\eta} = \frac{2.1 \times 1.2 \times 5.0 \times 10^{-3}}{1.8 \times 10^{-5}} \approx 7.0 \times 10^2$$

so the flow is laminar, since $R < 1000$. The flow would become turbulent at a speed above about:

$$\frac{R\eta}{\rho r} = \frac{1000 \times 1.8 \times 10^{-5}}{1.2 \times 5.0 \times 10^{-3}} = 3.0 \text{ m s}^{-1}$$

Nature of science

Understanding fluid flow

The study of fluid flow has made possible a large number of technological advances, including the design of more efficient and aerodynamic cars and aircraft, and new ways of measuring and understanding the flow of blood in arteries. Modelling the behaviour of winds and ocean currents has led to more accurate predictions of the weather. Fluid dynamics is an integral part of any study of stellar stability and stellar evolution. Research in fluid dynamics is ongoing and **turbulence** is still one of the great unsolved problems in the field, with many unanswered questions and formidable mathematical difficulties.

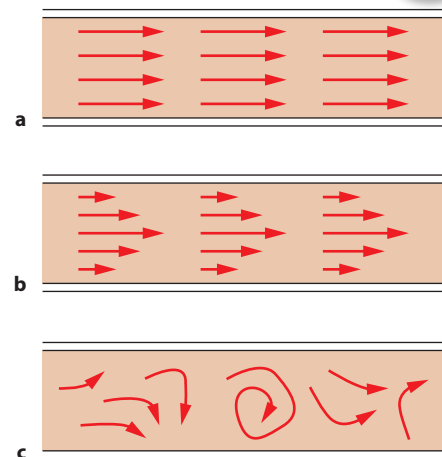
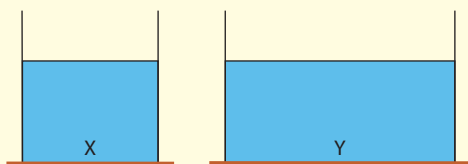


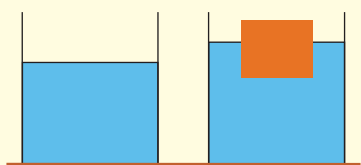
Figure B.59 a Non-viscous laminar flow; b viscous laminar flow; c turbulent flow.

? Test yourself

- 38 Two containers are filled with the same liquid to the same level. One container has double the cross-sectional area of the other. Compare the pressure at points X and Y.

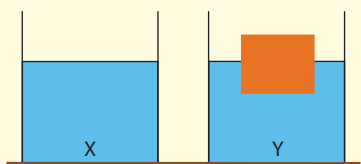


- 39 The diagram below shows two identical beakers containing equal amounts of water. In the second beaker, a piece of wood is floating in the water.

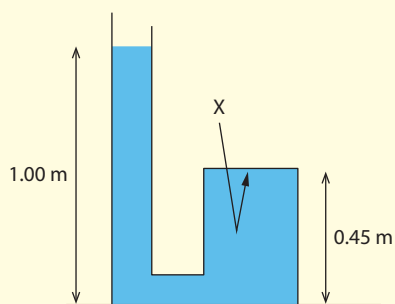


The beakers are weighed. Determine

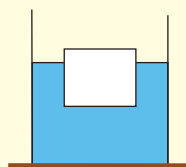
- which beaker, if any, is heavier
 - whether the pressure at the bottom of the two containers is the same.
- 40 The diagram below shows two identical beakers, filled with water to the same level. In the second beaker, a piece of wood is floating in the water.



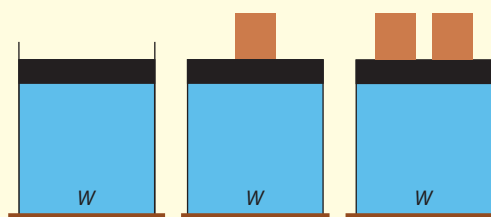
- The beakers are weighed. Determine which beaker, if any, is heavier.
 - Compare the pressure at points X and Y.
- 41 The container below is filled with water (density $1.0 \times 10^3 \text{ kg m}^{-3}$). Determine the pressure at point X. The liquid has density 13600 kg m^{-3} .



- 42 An ice cube floats in water. Explain why, after the ice cube melts, the level of the water in the container will be the same.

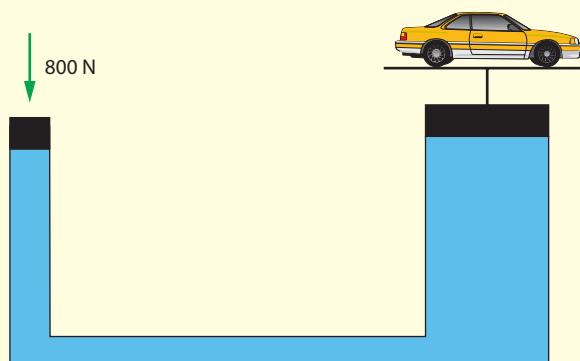


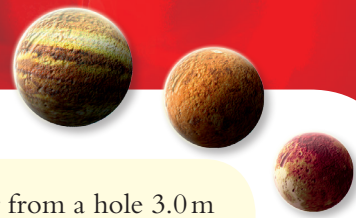
- 43 A liquid is enclosed in a container with a piston of negligible mass. Atmospheric pressure is p_0 .



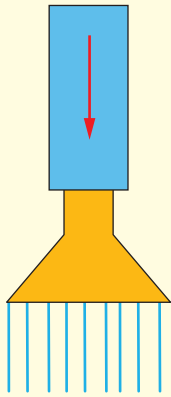
A block of weight W is placed on the piston, of area A . The weight is then doubled. Compare the pressures at the same depth h in each case.

- 44 A hollow sphere floats exactly half submerged in water of density $1.0 \times 10^3 \text{ kg m}^{-3}$. The outer radius of the sphere is 15 cm and the inner radius is 14 cm. Calculate the density of the material of the sphere. (The volume of a sphere of radius R is $V = \frac{4\pi R^3}{3}$.)
- 45 In the hydraulic pump below, a car of mass 1400 kg is to be lifted by applying a force of 800 N on a piston of diameter d . The diameter of the piston where the car is placed is 1.8 m. Calculate d .

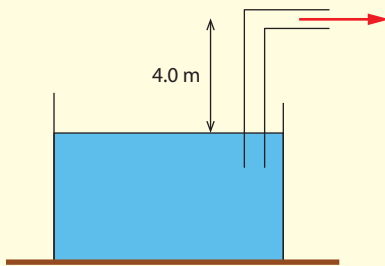




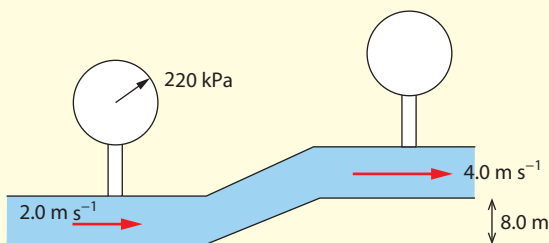
- 46 Water comes out of a tap of cross-sectional area 1.4 cm^2 . After falling a vertical distance of 5.0 cm , the cross-sectional area of the water column has been reduced to 0.60 cm^2 . Calculate the volume of water per second delivered by the tap.
- 47 In a shower, water enters the shower head through a tube of diameter 1.2 cm with a speed of 1.1 ms^{-1} . The shower head has 30 small holes, each of diameter 0.20 cm . Calculate the speed with which the water exits one of these holes.



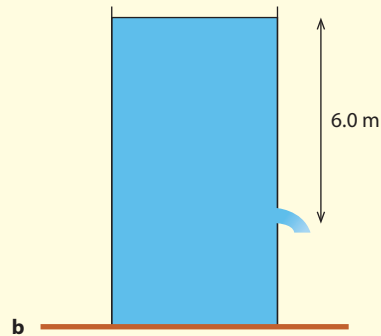
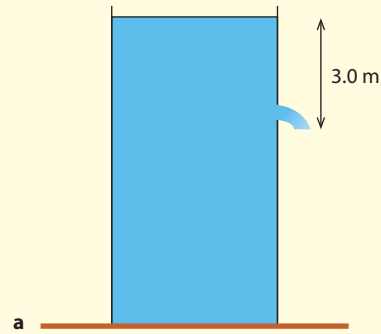
- 48 Water of density $1.0 \times 10^3 \text{ kgm}^{-3}$ is pumped out of a tank through a hose of radius 1.2 cm . The water in the hose has a constant speed of 3.8 ms^{-1} . The water is raised a vertical distance of 4.0 m before being ejected into the surroundings. Estimate the power of the pump.



- 49 Oil of density 850 kgm^{-3} flows in the pipe shown below. Calculate the pressure shown by the gauge on the upper side of the pipe.

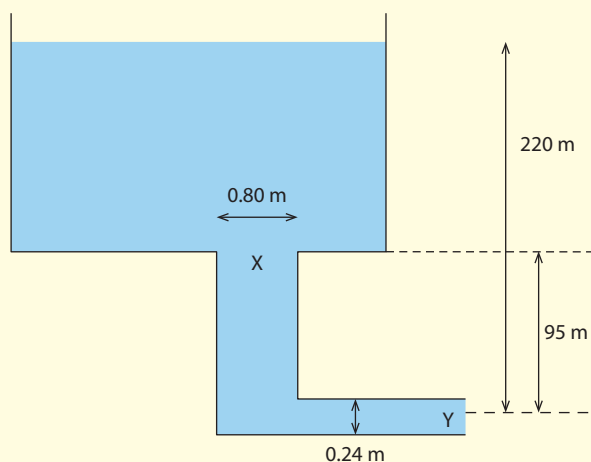


- 50 Water leaks out of a container from a hole 3.0 m below the free water surface. The container is large enough that no appreciable change occurs in the water level in the container.

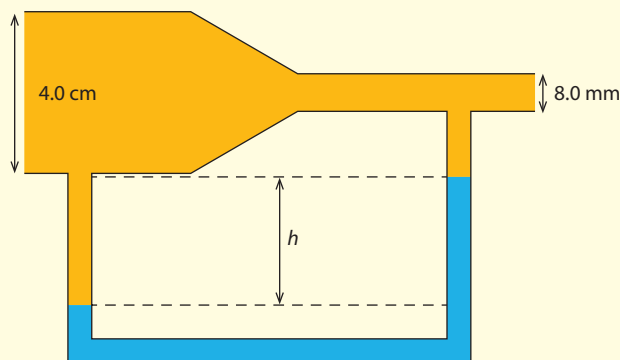


- a Calculate the speed at which the water exits the container.
- b A second hole is made at a depth of 6.0 m , at 3.0 m from the ground. Determine the ratio of the ranges (i.e. the horizontal distance travelled) of the two columns of water as they land on the ground.
- 51 Water exits a tank horizontally from a hole at a depth d . The water level in the tank is H and may be considered to be constant. Determine d in terms of H so that the water lands on the ground at the largest possible distance from the base of the tank.

- 52 In the diagram below (not drawn to scale), the exit at Y is closed with a tap so that no water flows.



- a Calculate the pressure at X and Y.
 b The exit at Y is opened so that Y is exposed to atmospheric pressure. Calculate:
 i the speed with which the water leaves the exit Y
 ii the pressure at X and at Y.
- 53 Determine the height h of the mercury in the Venturi tube in the diagram below. The fluid flowing is air (density 1.2 kg m^{-3}), at a volume flow rate of $1800 \text{ cm}^3 \text{ s}^{-1}$. The density of mercury is $13\,600 \text{ kg m}^{-3}$.



- 54 An aircraft is flying at an altitude where the air density is 0.35 kg m^{-3} . The pressure of the air outside the aircraft is $12\,000 \text{ Pa}$ lower than the pressure at the same altitude in static air in a Pitot-Prandtl tube. Calculate the speed of the aircraft.
- 55 Oil of density 850 kg m^{-3} and viscosity coefficient 0.01 Pa s flows in a pipeline of diameter 0.80 m . The flow rate of the oil in the pipeline is $0.52 \text{ m}^3 \text{ s}^{-1}$. Determine whether the flow is turbulent or laminar.
- 56 On a windy day, air flows in between tall buildings in a city. By making suitable estimates, determine whether the flow is turbulent or laminar.
- 57 An oil droplet (of density 870 kg m^{-3}) is balanced in between two oppositely charged, parallel capacitor plates. The droplet has a positive charge and the electric force on the droplet is directed vertically upwards. The electric field between the plates is $1.25 \times 10^5 \text{ N C}^{-1}$. When the field is turned off, the droplet falls and soon reaches a terminal velocity of $4.11 \times 10^{-4} \text{ m s}^{-1}$. Determine the charge on the droplet in terms of the electronic charge e . (The viscosity of air is $1.82 \times 10^{-5} \text{ Pa s}$ and you may neglect buoyancy effects.)

B4 Forced vibrations and resonance (HL)

In this section we examine the effect of resistance forces on simple harmonic oscillations. The effect of these forces is that the oscillations will eventually stop and the energy of the system will be dissipated, mainly as thermal energy in the environment and the system itself. We also consider the effect on an oscillating system of an externally applied periodic force. This leads to the interesting phenomenon of **resonance**.

B4.1 Oscillations and damping

In Topic 4 we saw examples of systems that are capable of oscillations or vibrations. Those were **free** oscillations: that is, oscillations without energy loss and without externally applied forces. In this case the amplitude of the oscillations stays constant (Figure B.60). In this section we will examine the effects of the presence of such forces.

The term **damping** refers here to the loss of energy in an oscillating system. Our discussion refers to frictional/resistance forces that are proportional to speed. We may distinguish three types of damping. In **light damping**, the amplitude of oscillation decreases slowly with time and eventually becomes zero; the oscillations stop (Figure B.61). Oscillations under light damping are also called **underdamped** oscillations. The amplitude decreases exponentially (dashed blue line).

The energy of the system also decreases exponentially (Figure B.62).

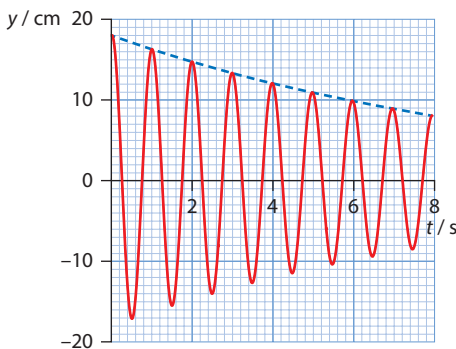


Figure B.61 Damped oscillations: the amplitude decreases exponentially with time.

How quickly will the oscillations in an underdamped system die out? A dimensionless number called the **Q factor** of the system answers this question. The Q factor is defined as:

$$Q = 2\pi \frac{\text{energy stored in a cycle}}{\text{energy dissipated in a cycle}}$$

Remember that the stored energy is proportional to the square of the amplitude. In Figure B.61, the initial amplitude is 18.0 cm and after one full oscillation it is about 16.3 cm. This means that:

$$Q = 2\pi \frac{18.0^2}{18.0^2 - 16.3^2} \approx 35$$

A high Q value means that the system will perform many oscillations before the oscillations die out. Roughly, Q is the number of oscillations before the system stops.

Learning objectives

- Understand the concept of natural frequency of vibration.
- Work with damping and the Q factor.
- Work with forced oscillations.
- Appreciate the concept of resonance.

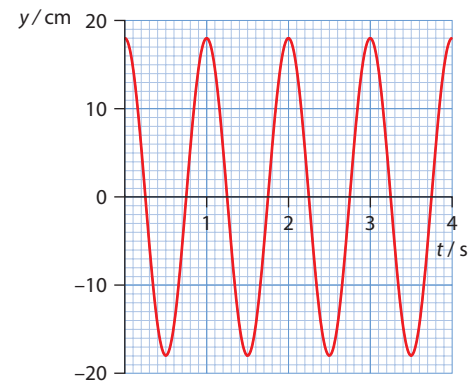


Figure B.60 Undamped oscillations: the amplitude is constant.

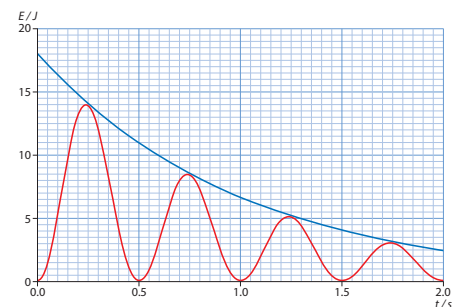


Figure B.62 In damped motion the total energy of a system decreases exponentially (blue curve). The red curve shows the kinetic energy of the system. The period of oscillations is 1.0 s.

The expression for Q may be written in an equivalent way: let P be the power loss of the system. Then in one cycle the energy dissipated is $E_{\text{dissipated}} = PT$, where T is the period of oscillations. Thus:

$$\begin{aligned} Q &= 2\pi \frac{E_{\text{stored}}}{E_{\text{dissipated}}} \\ &= 2\pi \frac{E_{\text{stored}}}{PT} \\ &= 2\pi \frac{1}{T} \frac{E_{\text{stored}}}{P} \\ &= 2\pi f \frac{E_{\text{stored}}}{P} \end{aligned}$$

where f is the frequency at which the system oscillates.

If the amount of damping is very large, we speak of **overdamped** motion. In this case the system returns to its equilibrium position after a very long time **without performing oscillations** (blue curve in Figure B.63).

For a particular amount of damping we have the case of **critically damped** motion, in which the system returns to its equilibrium without any oscillations but does so in the shortest possible time (purple curve in Figure B.63).

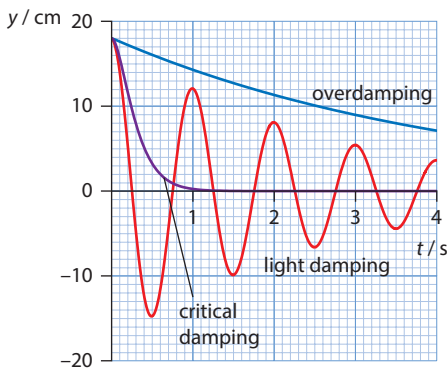


Figure B.63 Lightly damped oscillations (red curve), overdamped motion (blue curve) and critically damped motion (purple curve).

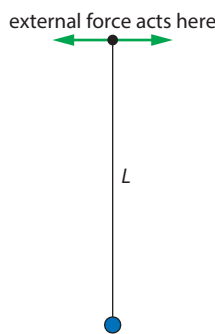


Figure B.64 Forced oscillations: a periodic force makes the point of support vibrate.

B4.2 Forced oscillations and resonance

We will now examine qualitatively the effect of an externally applied force F on a system that is free to oscillate with frequency f_0 . The force F will be assumed to vary periodically with time with a frequency (the driving frequency) f_D – for example, $F = F_0 \cos 2\pi f_D t$. The oscillations in this case are called **forced** or **driven** oscillations. As an example consider a pendulum of length L that hangs vertically, as in Figure B.64. The point of support is made to oscillate with some frequency f_D . How does the pendulum react to this force?

In general, once the external force is applied, the system eventually starts oscillating at the driving frequency f_D . This is true whether the system is initially at rest or initially oscillating at its own frequency. In the absence of friction, the amplitude of the oscillations approaches infinity as the driving frequency approaches the **natural frequency** of the oscillating system (Figure B.65). Of course an infinite amplitude is impossible: we must also take damping into account.

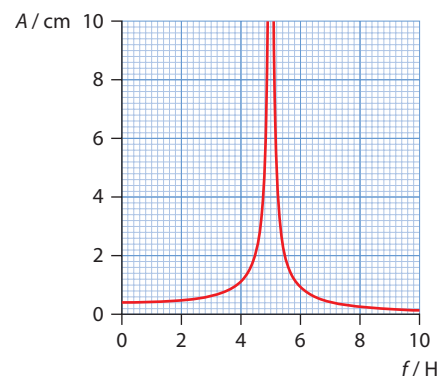


Figure B.65 Variation with driving frequency of the amplitude of simple harmonic motion when an undamped system is driven by an external periodic force.

The amplitude of the oscillations will depend on the relationship between f_D and f_0 , and the amount of damping. We might expect that, because the system prefers to oscillate at its own natural frequency, when the external force has the same frequency as the natural frequency, large oscillations will take place. A detailed analysis produces the graph in Figure B.66, showing how the amplitude of oscillation of a system with natural frequency f_0 varies when it is subjected to a periodic force of frequency f_D . The degree of damping increases as we move from the top curve down.

The general features of the graph in Figure B.66 are as follows:

- For small damping, the peak of the curve occurs at the natural frequency of the system, f_0 .
- The lower the damping, the higher and narrower the curve.
- As the amount of damping increases, the peak shifts to lower frequencies and becomes wider.
- At very low frequencies, the amplitude is essentially constant.

If f_D is very different from f_0 , the amplitude of oscillation will be small. On the other hand, if f_D is approximately the same as f_0 and the degree of damping is small, the resulting driven oscillations will have large amplitude. The largest amplitude is obtained when f_D is equal to f_0 , in which case we say that the system is in resonance.

The state in which the frequency of an externally applied periodic force equals the natural frequency of a system is called **resonance**. This results in oscillations with large amplitude.

Resonance can be disastrous: we do not want an aircraft wing to resonate, nor is it good for a building or a bridge to be set into resonance by an earthquake or the wind (Figure B.67).



Figure B.67 The Tacoma Narrows Bridge in Washington state in the USA, oscillating before collapsing in 1940, a victim of resonance caused by high winds.

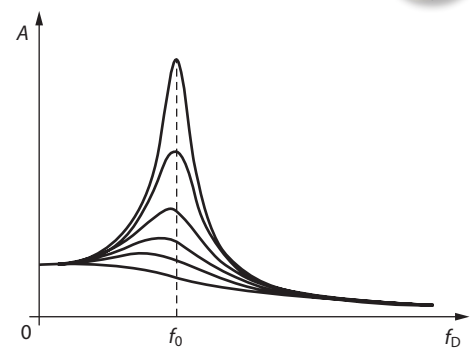


Figure B.66 Variation of amplitude with driving frequency and degree of damping when a damped system is driven by an external periodic force.

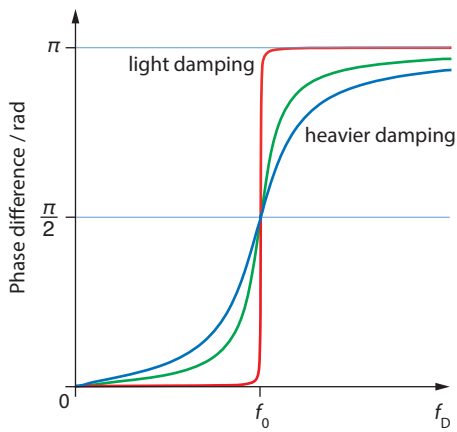


Figure B.68 Phase difference as a function of the external driving frequency.

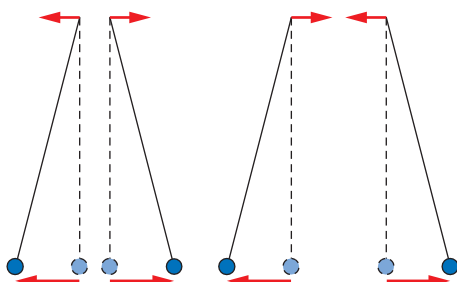


Figure B.69 **a** The system and the driver oscillate in phase. **b** The system and the driver oscillate out of phase.

Resonance can be irritating if the car in which you drive is set into resonance by bumps on the road or by a poorly tuned engine. But resonance can also be a good thing: it is used in a microwave oven to warm food, and your radio uses resonance to tune into one specific station and not another. Another useful example of electrical resonance is the quartz oscillator, a quartz crystal that can be made to vibrate at a specific frequency.

In the case of driven oscillations there is a phase difference between the displacement of the system and the displacement of the driver. This phase difference depends on the relation between the external driving frequency and the natural frequency. The relationship is shown in Figure B.68.

For very light damping, the phase difference is zero when $f < f_0$: in this case the system and the driver oscillate in phase (Figure B.69a). When $f > f_0$, the phase difference is π : the system and the driver are completely out of phase (Figure B.69b). There is a phase difference of $\frac{\pi}{2}$ when $f = f_0$.

Nature of science

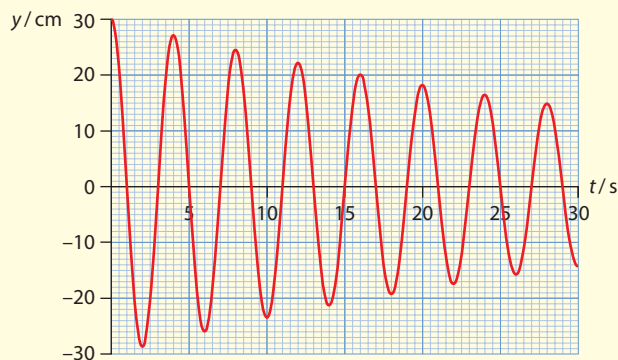
Risk assessment

The phenomenon of resonance is ubiquitous in physics and engineering. Microwave ovens emit electromagnetic radiation in the microwave region of wavelengths that match the vibration frequencies of water molecules so they can be absorbed and warm food. In the medical technique known as MRI, the patient is exposed to a strong magnetic field and resonance is used to force absorption of radio frequency radiation by protons in the patient's body. This technique provides detailed images of body organs and cellular functions. Just as with buildings and bridges, detailed studies are necessary to avoid the possibility of unwanted resonances creating catastrophic large-amplitude oscillations.

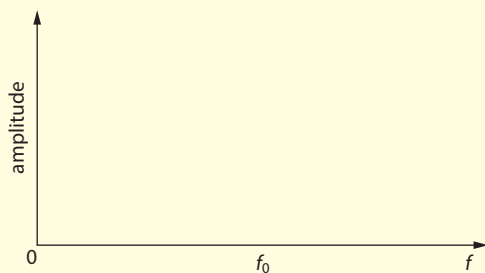


? Test yourself

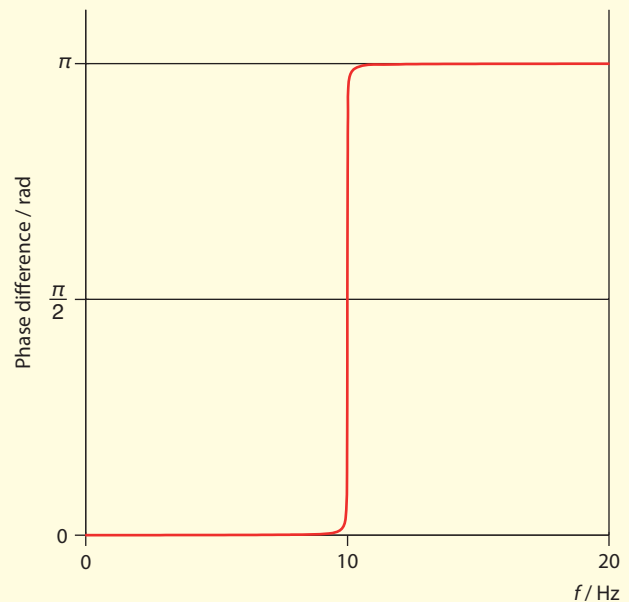
- 58 Distinguish between **free** and **driven** oscillations.
- 59 Distinguish between **overdamped** and **critically damped** oscillations.
- 60 In the context of oscillations, state what is meant by **damping**.
- 61 State what is meant by **resonance**.
- 62 The graph below shows the variation of the displacement of a system performing oscillations.



- a Determine the frequency of the oscillations.
- b Calculate the Q factor of the system.
- c Draw a sketch graph to show the variation with time of the energy of the system.
- d The amount of damping is increased. State and explain the effect of this on the value of Q .
- 63 a State the conditions that must be satisfied for simple harmonic oscillations to take place.
- b Draw a graph of displacement versus time for a body undergoing lightly damped simple harmonic oscillations.
- c A system has natural frequency f_0 . A periodic force of variable frequency f acts on the system. On a copy of the figure below, draw graphs to show the variation with f of the amplitude of oscillation of the system for **i** light and **ii** heavy damping.



- 64 A body performs simple harmonic oscillations at the end of a vertical spring. The diagram below shows the variation of the phase difference between the displacement of the body and the displacement of a driver that drives the system with driver frequency f . The system is very lightly damped.

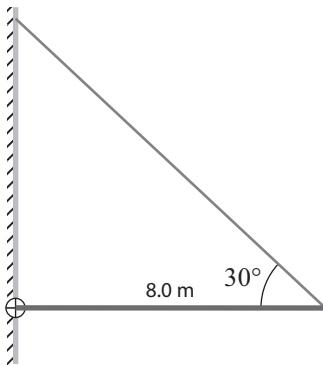


- a State the natural frequency of oscillation of the system.
- b Copy the figure and on the same axes draw a sketch graph to show how this graph changes, if at all, when the damping is increased.
- c The frequency of the driver is 15 Hz. At a particular time the driver forces the top end of the spring to move to the right. State and explain the direction of motion of the mass at the end of the spring at that instant.

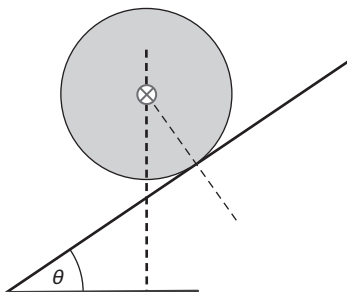
Exam-style questions

Note: You may use the textbook to find any moments of inertia you may require.

- 1 A uniform plank of length 8.0 m and weight 1500 N is supported horizontally by a cable attached to a vertical wall. (You may use the textbook to find any moments of inertia you may require.) The cable makes an angle of 30° with the horizontal rod.



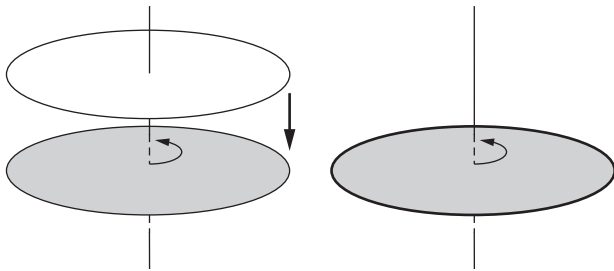
- a Calculate:
- the tension in the cable [3]
 - the magnitude and direction of the force exerted by the wall on the rod. [3]
- b A worker of mass 85 kg can walk anywhere on the rod without fear of the cable breaking. Determine the minimum breaking tension of the cable. [3]
- 2 A cylinder of mass $M = 12.0$ kg and radius $R = 0.20$ m rolls down an inclined plane without slipping.



- a Make a copy of the figure, and on it draw arrows to represent the forces acting on the cylinder as it rolls. [3]
- b
- Show that the linear acceleration of the centre of mass of the cylinder is $a = \frac{2}{3}g \sin \theta$. [3]
 - Determine the frictional force acting on the cylinder for $\theta = 30^\circ$. [1]
- c Calculate the rate of change of the angular momentum of the cylinder as it rolls for $\theta = 30^\circ$. [2]

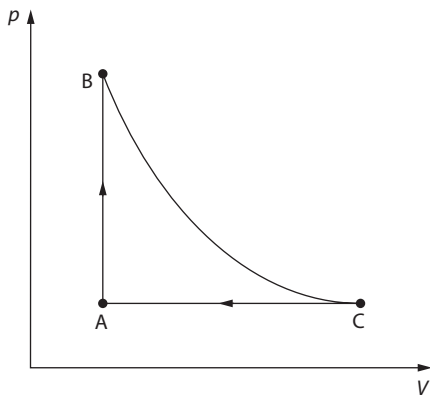


- 3 A horizontal disc rotates about a vertical axis through its centre of mass. The mass of the disc is 4.00 kg and its radius is 0.300 m. The disc rotates with an angular velocity of 42.0 rad s^{-1} . A ring of mass 2.00 kg and radius 0.300 m falls vertically and lands on top of the disc, as shown below. As the ring lands, it slides a bit on the disc and eventually the disc and ring rotate with the same angular velocity.



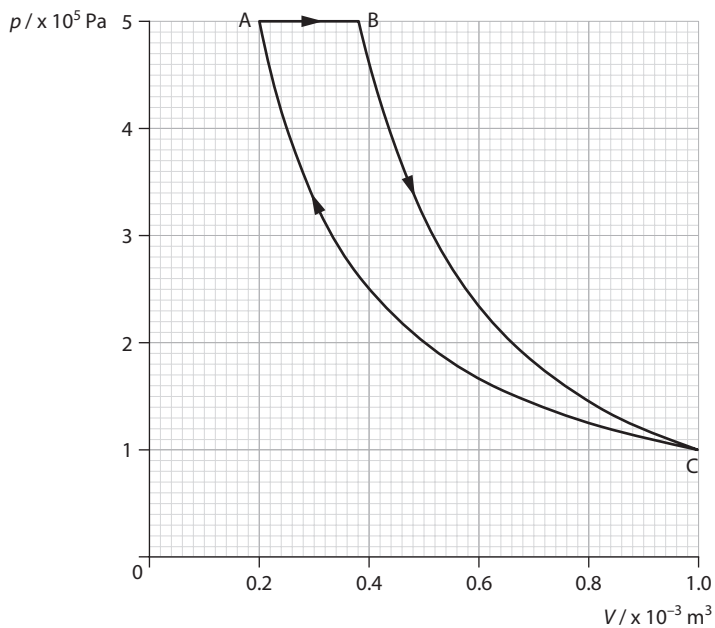
- a
- i Explain why the angular momentum of the system is the same before and after the ring lands. [2]
 - ii Calculate the final angular velocity of the disc–ring system. [3]
 - iii Determine the kinetic energy lost as a result of the ring landing on the disc. [3]
- b It took 3.00 s for the ring to start rotating with the same angular velocity as the disc.
- i Determine the average angular acceleration experienced by the ring during this time. [1]
 - ii Calculate the number of revolutions made by the disc during the 3.00 s. [2]
 - iii Show that the torque that accelerated the ring to its final constant angular velocity was 1.26 N m. [2]
 - iv State and explain, without further calculation, the magnitude of the torque that decelerated the disc. [2]
- c Calculate the average power developed by the torque in b iii in accelerating the ring. [2]
- 4 A heat engine has 1.00 moles of an ideal gas as its working substance and undergoes the cycle shown below. BC is an adiabatic. The following data are available:

$$T_A = 3.00 \times 10^2 \text{ K}, p_A = 2.00 \times 10^5 \text{ Pa}, T_B = 6.00 \times 10^2 \text{ K}$$

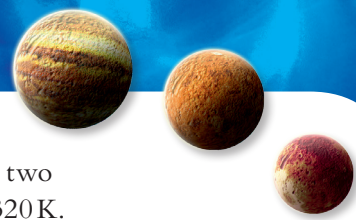


- a Calculate:
- i the pressure at B [1]
 - ii the temperature at C [2]
 - iii the volume at A and at C. [2]
- b Determine:
- i the change in internal energy from A to B [2]
 - ii the energy removed from the gas. [2]
 - iii the efficiency of the cycle [2]
- c State a version of the second law of thermodynamics in a way that directly applies to this engine. [2]

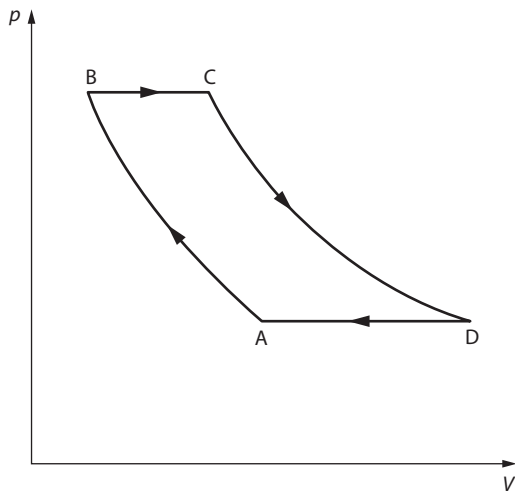
- 5 The graph shows a thermodynamic cycle in which an ideal gas expands from A to B to C and is then compressed back to A. BC is an adiabatic curve and CA is an isothermal curve. The volume at B is $0.38 \times 10^{-3} \text{ m}^3$.



- a i State what is meant by an **adiabatic curve**. [1]
 ii Explain why, in an adiabatic expansion of an ideal gas, temperature decreases. [2]
- b Justify why CA is isothermal. [3]
- c The temperature of the gas at A is 300 K.
 i Calculate the temperature at B and at C. [3]
 ii Determine the number of moles of the gas. [2]
- d The work done from C to A is 160 J. Calculate:
 i the energy transferred out of the gas [2]
 ii the energy transferred into the gas [2]
 iii the work done from B to C [2]
 iv the efficiency of the cycle. [1]

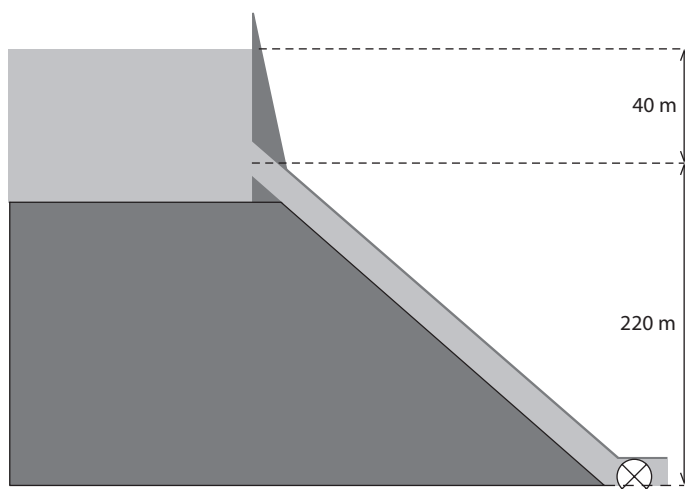


- HL 6** The graph shows the Brayton cycle (not drawn to scale), which consists of two isobaric and two adiabatic curves. The state at A has pressure 2.0×10^5 Pa, volume 0.40 m^3 and temperature 320 K . The pressure at B is 2.0×10^6 Pa.

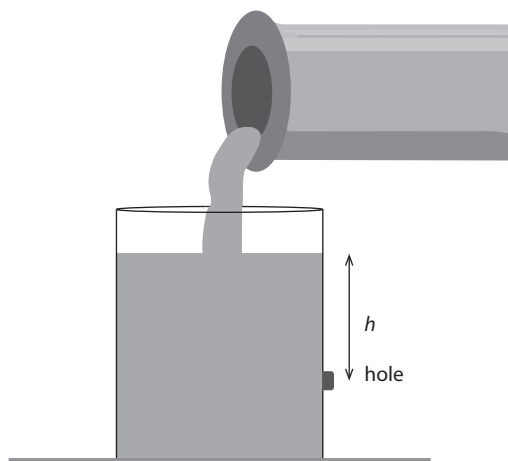


- a** Show that, along an adiabatic curve, $\frac{T^5}{p^2} = \text{constant}$. [3]
- b** Calculate:
- the temperature at B [2]
 - the volume at B. [2]
- c** Show that the change in internal energy from A to B is 0.18 MJ . [3]
- d** Using your answer to **c**, determine the work done from A to B. [2]
- HL 7 a** Show that the pressure at a depth h below the free surface of a liquid of density ρ is given by $p = p_0 + \rho gh$, where p_0 is atmospheric pressure. [3]
- b** Suggest what, if anything, will happen to the pressure at a depth h below the free surface of the liquid in a container, if the container:
- is allowed to fall freely under gravity [1]
 - is accelerated upwards with acceleration a . [1]
- c** State **Archimedes' principle**. [1]
- d** A block of wood floats in water with 75% of its volume submerged. The same block when floating in oil has 82% of its volume submerged. The density of water is 1000 kg m^{-3} . Calculate:
- the density of the wood [2]
 - the density of the oil. [2]

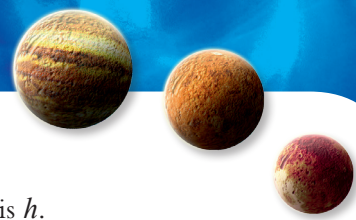
- HL 8** The diagram shows a hydroelectric power plant in which water from a large reservoir is allowed to flow down a long outlet pipe and, eventually, through a turbine. The radius of the outlet pipe where it leaves the reservoir is 65 cm, and it tapers down to 25 cm at the turbine. You may assume that the reservoir is large enough that there is no appreciable change in the water level of the reservoir as water flows out.



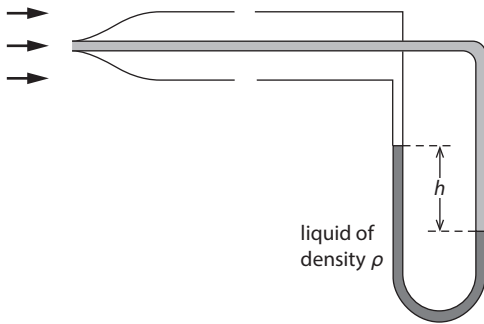
- a i** Calculate the speed of the water at the turbine. [2]
ii State two assumptions you have made in this calculation. [2]
- b** Estimate the flow rate of the water through the turbine. [2]
- c** Calculate the water pressure at the upper end of the outlet pipe:
i when the water is static (i.e. not flowing) [2]
ii when the water is flowing. [3]
- d** In another, smaller reservoir, water is constantly being pumped into the reservoir at a rate of $0.40 \text{ m}^3 \text{ s}^{-1}$.



A hole of radius 3.0 cm is to be drilled at a depth h below the surface of the water such that, when the water flows out, the water level in the reservoir stays the same. Determine the depth h . [3]

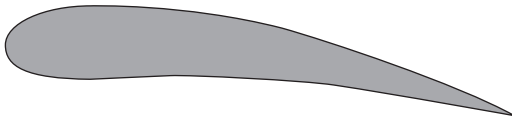


- HL 9** A Pitot–Prandtl tube is used to measure the speed of an aircraft. The liquid in the manometer has density ρ and the difference in the levels of the liquid in the two columns is h .



- a Explain why the liquid in the right-hand column of the manometer is lower than that in the left-hand column. [2]
- b Show that the flow speed v is given by $v = \sqrt{\frac{2\rho g h}{\rho_{\text{air}}}}$ [3]
- c This Pitot tube uses a liquid of density $\rho = 920 \text{ kg m}^{-3}$. The density of air is 1.20 kg m^{-3} and the difference h in liquid levels is 0.25 m . Estimate the speed of the aircraft. [2]

- HL 10** The diagram shows an aerofoil of **total** surface area 16 m^2 . Air (of density 1.20 kg m^{-3}) flows with a speed of 85 m s^{-1} across the upper surface of the aerofoil and a speed of 58 m s^{-1} across the lower surface.



- a On a copy of the diagram, draw streamlines around the aerofoil. [2]
- b i Calculate the lifting force on the aerofoil. [2]
 ii State one assumption you made in your calculation of the lifting force. [1]
- c The weight of the aerofoil is 3.0 kN . The aerofoil is attached to the fuselage of an aircraft. Estimate the force that the aerofoil exerts on the fuselage. [2]
- d If the aerofoil's angle relative to the horizontal is increased, the flow of air past it may become turbulent.
 i State what is meant by turbulent flow. [1]
 ii Indicate on your copy of the diagram the most likely position around the aerofoil where turbulence may set in. [1]
 iii Suggest the effect of turbulence on the lifting force on the aerofoil. [1]

- HL 11 a** Distinguish between **damped** and **undamped** oscillations. [2]
- b** The graph shows the variation with time of the displacement of an oscillating system.

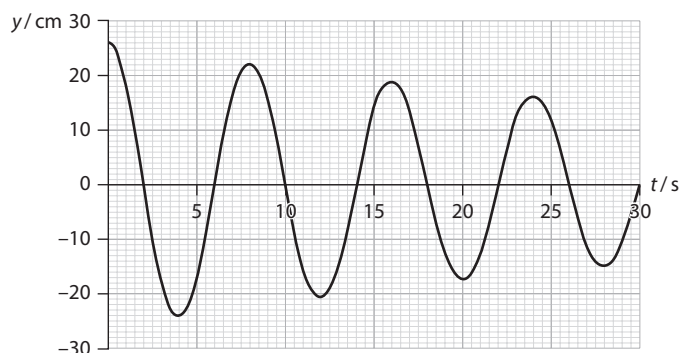
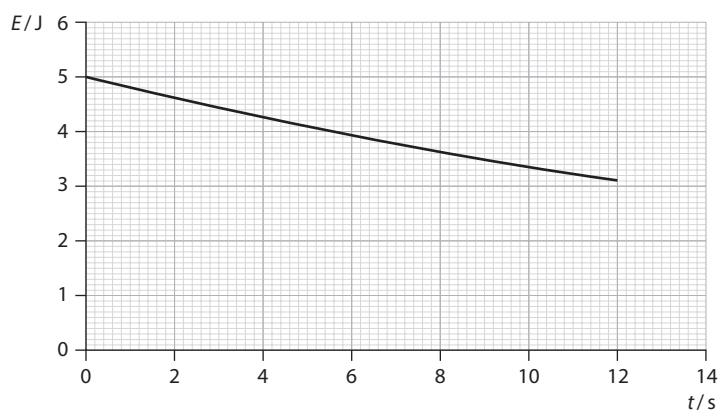


Figure B.110 For Exam-style question 11b.

Use the diagram to determine:

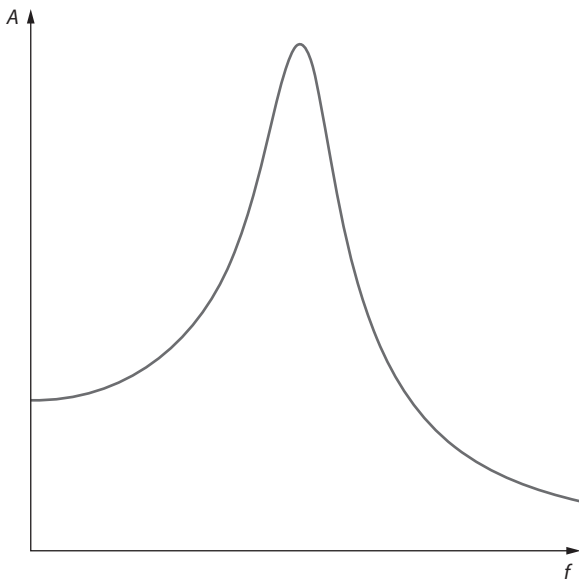
- i** the period of oscillation [1]
- ii** the Q factor of the system. [3]
- c i** On a copy of this diagram, draw a graph to show the variation of the displacement when the damping is increased. [2]
- ii** State the effect, if any, of the increased damping on Q . [1]
- d** The graph shows the variation with time of the energy of another oscillating system.



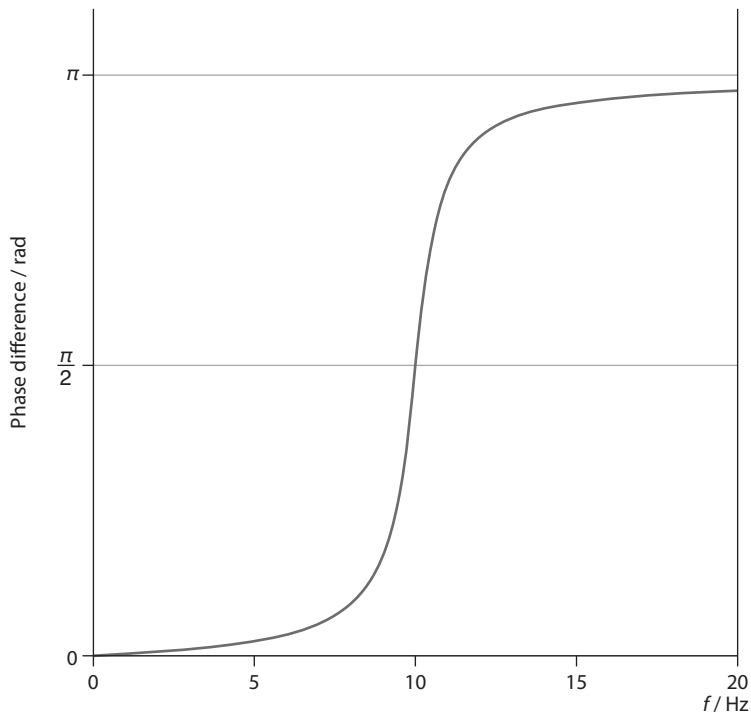
The system oscillates with a period of 2.0s. Estimate the Q factor of this system. [2]



- HL 12** The graph shows the variation of the amplitude of an oscillating system with the frequency of a periodic force acting on the system. The system is lightly damped.



- a By reference to the diagram, outline what is meant by resonance. [3]
- b On a copy of this diagram, sketch the variation of amplitude with frequency when the amount of damping is increased. [2]
- c The graph shows how the phase difference between the displacements of the system and the driver varies with the frequency of the driving force, for light damping.



- i On a copy of this diagram, sketch a graph to show how the phase varies when the damping is increased. [2]
- ii State the resonant frequency of this system. [1]

Option C Imaging

C1 Introduction to imaging

In this section we discuss the formation of images by lenses and mirrors. We will learn how to construct images graphically as well as algebraically. The section closes with a discussion of two types of lens defect: spherical and chromatic aberration.

C1.1 Lenses

The passage of light rays through lenses is determined by the law of refraction. A ray of light that enters a lens will, in general, deviate from its original path according to **Snell's law** of refraction.

The extent of deviation depends on the index of refraction of the glass making up the lens, the radii of the two spherical surfaces making up the lens, and the angle of incidence of the ray.

We will make the approximation that the lens is always very **thin**, which allows for simplifications. The two sides of the lens need not have the same curvature, and may be convex, concave or planar. Various types of lens are illustrated in Figure C.1.

The straight line that goes through the centre of the lens at right angles to the lens surface is known as the **principal axis** of the lens.

C1.2 Converging lenses

Lenses that are thicker at the centre than at the edges are **converging lenses**, which means that, upon going through the lens, a ray of light changes its direction towards the axis of the lens (see Figure C.2a).

The straight line at right angles to the lens surface and through its centre is called the principal axis of the lens. A beam of rays parallel to the principal axis will, upon refraction through the lens, pass through

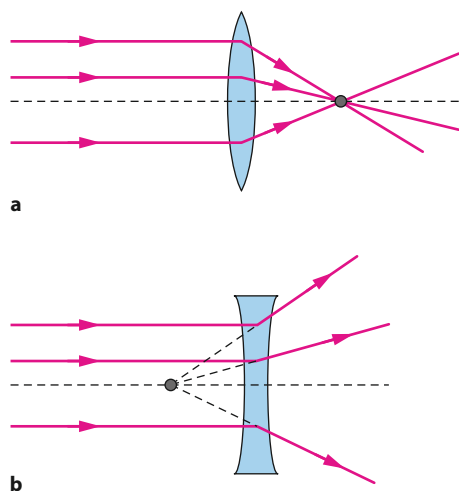


Figure C.2 a A converging lens. b A diverging lens. A beam of rays parallel to the principal axis converges towards the principal axis in the case of a converging lens but diverges from it in the case of a diverging lens.

Learning objectives

- Work with thin converging and diverging lenses.
- Work with concave and convex mirrors.
- Solve problems with ray diagrams, both graphically and algebraically.
- Understand the difference between real and virtual images.
- Calculate linear and angular magnifications.
- Describe spherical and chromatic aberrations.

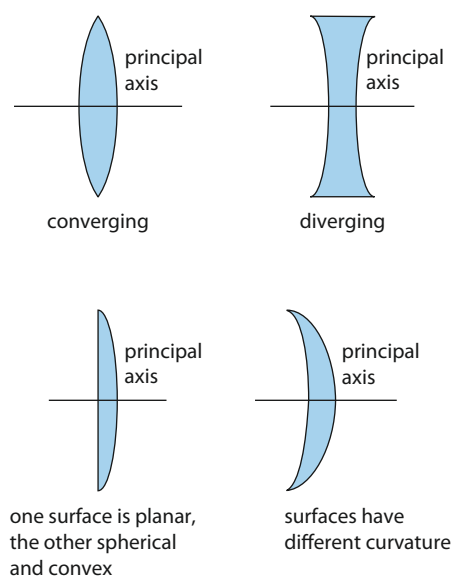


Figure C.1 Various types of lenses.

the same point on the principal axis on the other side of the lens (see Figure C.3a).

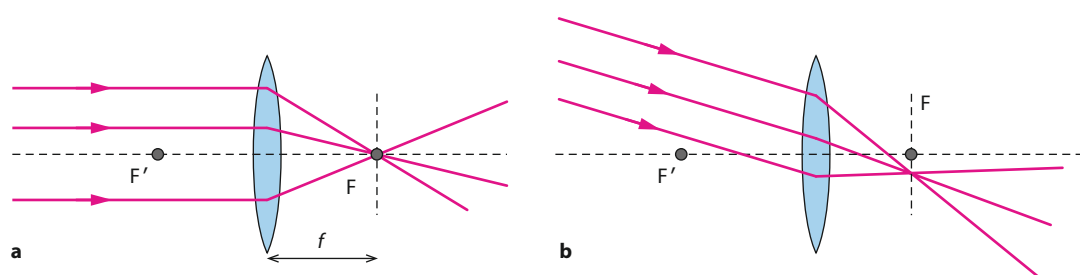


Figure C.3 **a** Upon refraction, rays that are parallel to one another and to the principal axis pass through the focal point of the lens, a point on the principal axis. **b** If the rays are not parallel to the principal axis, they will go through a common point that is in the same plane as the focal point.

Exam tip

Optometrists usually use the inverse of the focal length – called the power, P – to specify a lens:

$$P = \frac{1}{f}$$

If the focal length of a lens is expressed in metres, its power is expressed in dioptres, D ($D = 1 \text{ m}^{-1}$).

For example, a lens with a focal length $f = 25 \text{ cm}$ has a power of $P = \frac{1}{0.25} = 4.0 \text{ D}$

Rays that are parallel to the principal axis will, upon refraction, pass through a point on the principal axis called the **focal point**. The distance of the focal point from the centre of the lens is called the **focal length**, denoted f .

If a parallel beam of rays is not parallel to the axis, the rays will again go through a single point. This point and the focal point of the lens are in the same vertical plane (see Figure C.3b).

(The point on the other side of the lens at a distance f from the lens is also a focal point. A ray parallel to the principal axis and entering the lens from right to left will pass through F' .)

We now know how one set of rays will refract through the lens. Let us call a ray parallel to the principal axis ‘standard ray 1’.

Another ray whose refraction through the lens is easy to describe passes through the left focal point of the lens. It then emerges parallel to the principal axis on the other side of the lens, as shown in Figure C.4. We may call such a ray ‘standard ray 2’.

A third light ray whose behaviour we know something about is directed at the centre of the lens. This ray will go through undeflected, as shown in Figure C.5a. We may call such a ray ‘standard ray 3’.

The reason for this behaviour is that near the midpoint of the lens the two lens surfaces are almost parallel. A ray of light going through glass with two parallel surfaces is shown in Figure C.5b. The ray simply gets shifted parallel to itself. The amount of the parallel shift is proportional to the width of the glass block, which is the thickness of the lens. Since we are making the approximation of a very thin lens, this displacement is negligible.

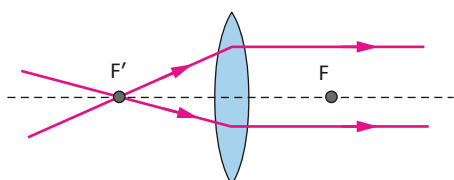


Figure C.4 A ray passing through the focal point emerges parallel to the principal axis.

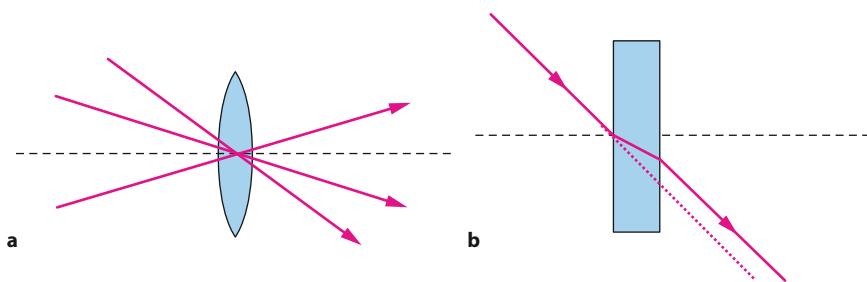


Figure C.5 **a** A ray going through the centre of the lens is undeflected. **b** A ray entering a glass plate is shifted parallel to itself by an amount proportional to the thickness of the plate.

With the help of these three ‘standard rays’ we can find the image of any object placed in front of a converging lens. The standard rays are shown together in Figure C.6.

We will get different kinds of images depending on the distance of the object from the lens. The distances of the object and the image are measured from the centre of the lens.

The convention shown in Figure C.7 for the representation of a lens by a single line, rather than by its actual shape, is helpful in simplifying graphical solutions.

We start with Figure C.8, showing an object 12 cm from the lens. The object is 4.0 cm high. The focal length of the lens is 4.0 cm. We draw the three standard rays leaving the top of the object. These meet at point P on the other side of the lens, and that point is the image of the top of the object. Because the object is at right angles to the principal axis, the rest of the image is found by just drawing a vertical arrow from the principal axis to the point P. We observe that the distance of the image is about 6.0 cm and its height is about 2.0 cm. The image is inverted (upside down).

In our second example, the object is placed in between the lens and the focal point of the lens (Figure C.9).

Here the object distance is 4.0 cm and the focal length is 6.0 cm. We draw the standard rays. Standard ray 2 is awkward: how can we draw this ray passing through the focal point and then refracting through the lens? We do so by imagining a backwards extension that starts at the focal point.

The diagram shows that the three refracted rays do not cross on the other side of the lens. In fact they diverge, moving away from each other. But if we extend these refracted rays backwards, we see that their extensions meet, at point P. An observer on the right side of the lens, seeing the refracted rays, would think that they originated at P. The rest of the image is constructed by drawing a vertical arrow to P from the principal axis. The image is thus formed on the same side of the lens as the object, upright and larger. Its height is 9.0 cm and its distance is 12 cm.

There is an essential difference between the images in Figures C.8 and C.9. In Figure C.8, actual rays pass through the image. In Figure C.9 no rays originate from the image; only the mathematical extensions of the rays do. In the first case a screen placed in the image plane would show the image on the screen. A screen placed in the image plane in the second case would show nothing. The rays of light would increase the temperature at the position of the first image, but no such rise in temperature would occur in the second case.

The first image is called **real** and the second **virtual**.

A **real image** is formed by actual rays and can be projected on a screen. A **virtual image** is formed by extensions of rays and cannot be projected on a screen.

It is left as an exercise to draw the **ray diagram** for an object that is placed at a distance from the lens which is exactly equal to the focal length. You will find that the refracted rays are parallel. In this case neither the rays themselves nor their extensions meet. The image is said to form at infinity.

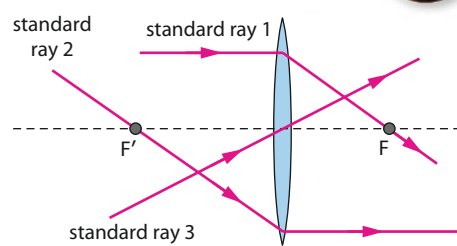


Figure C.6 Refraction of the three standard rays in a converging lens.

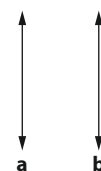


Figure C.7 Convention for representing **a** converging and **b** diverging lenses.

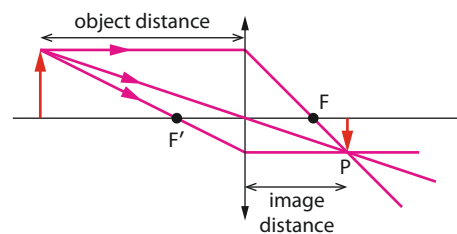


Figure C.8 Image formation in a converging lens. The image is real.

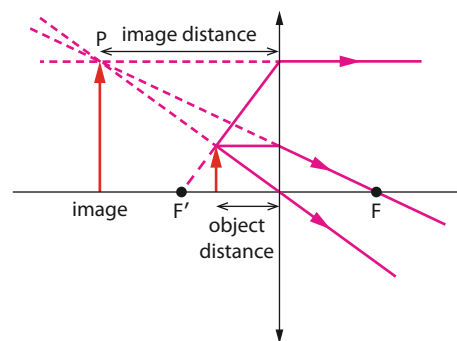


Figure C.9 Image formation in a converging lens. The image is virtual.

Exam tip

You can form the image using just two of the three standard rays, but the third ray provides a check on your drawing.

Worked example

C.1 Figure C.10 shows the image of an object in a lens. A ray of light leaves the top of the object. **a** On a copy of this diagram, draw a line to show how this ray refracts in the lens. **b** Draw another line to locate the focal point of the lens.

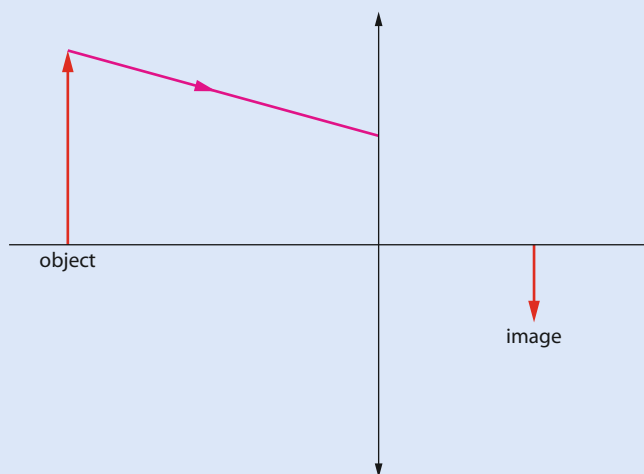


Figure C.10

Exam tip

In drawing ray diagrams it is best to represent lenses by straight lines rather than by their actual shape.

- a** Since the ray leaves the top of the object (at the arrow) it must go through the arrow in the image.
- b** Draw a ray from the arrow for the object parallel to the principal axis. This ray, when refracted, must also go through the arrow for the image. Where it crosses the principal axis is the focal point (Figure C.11).

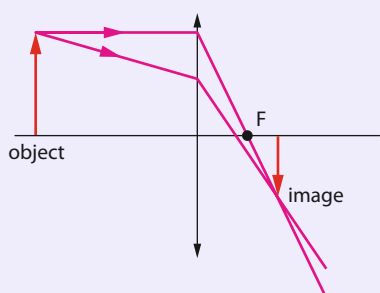


Figure C.11

The methods described above are graphical methods for finding an image. These are very useful because they allow us to ‘see the image being formed’. There is, however, also an algebraic method, which is faster. This uses an equation relating the object and image distances to the focal length of the lens.

We can derive this equation as follows. The object is placed in front of the lens, as shown in Figure C.12.

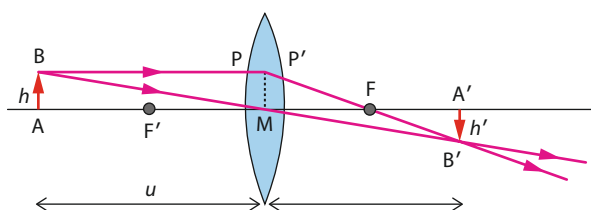


Figure C.12 The image of an object placed in front of a converging lens.



Let u be the distance of the object from the lens and v the distance of the image from the lens. Triangles ABM and A'B'M are similar (their angles are clearly equal), so

$$\frac{h}{u} = \frac{h'}{v} \Rightarrow \frac{h'}{h} = \frac{v}{u}$$

The lens is thin, so P and P' may be considered to be the same point. Then triangles MPF and A'B'F are similar, so

$$\frac{h}{f} = \frac{h'}{v-f} \Rightarrow \frac{h'}{h} = \frac{v-f}{f}$$

(note that MP = h). Combining the two equations gives

$$\frac{v}{u} = \frac{v-f}{f}$$

$$\Rightarrow vf = uv - uf$$

$$\Rightarrow vf + uf = uv \text{ (divide by } uvf)$$

$$\Rightarrow \frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

The last equation is known as the **thin-lens equation**, and may be used to obtain image distances.

To examine whether the image is larger or smaller than the object, we define the **linear magnification**, m , of the lens as the ratio of the image height to the object height:

$$m = \frac{\text{image height}}{\text{object height}}$$

Numerically, the linear magnification is $m = \frac{v}{u}$. It turns out to be convenient to introduce a minus sign, so we define linear magnification as

$$m = \frac{\text{image height}}{\text{object height}} = \frac{h'}{h} = -\frac{v}{u}$$

The usefulness of the minus sign will be appreciated in the following Worked examples.

The thin-lens equation and the magnification formula allow a complete determination of the image without a ray diagram. But to do this, a number of conventions must be followed:

- f is positive for a converging lens
- u is positive
- v is positive for real images (those formed on the other side of the lens from the object)
- v is negative for virtual images (those formed on the same side of the lens as the object)
- $m > 0$ means the image is upright
- $m < 0$ means the image is inverted
- $|m| > 1$ means the image is larger than the object
- $|m| < 1$ means the image is smaller than the object.

Worked examples

C.2 A converging lens has a focal length of 15 cm. An object is placed 60 cm from the lens. Determine the size of the image and the value of the magnification.

The object distance is $u = 60$ cm and the focal length is $f = 15$ cm. Thus,

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{15} - \frac{1}{60} = \frac{1}{20} \Rightarrow v = 20 \text{ cm}$$

The image is real (positive v) and is formed on the other side of the lens. The magnification is $m = \frac{20}{60} = \frac{1}{3}$. The negative sign in the magnification tells us that the image is inverted. The magnitude of the magnification is less than 1. The image is three times shorter than the object. (Construct a ray diagram for this example.)

C.3 An object is placed 15 cm in front of a converging lens of focal length 20 cm. Determine the size of the image and the value of the magnification.

Applying the lens equation, we have

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{20} - \frac{1}{15} = -\frac{1}{60} \Rightarrow v = -60 \text{ cm}$$

The image is virtual (negative v) and is formed on the same side of the lens as the object. The magnification is $m = -\left(-\frac{60}{15}\right) = +3$. Thus the image is three times taller than the object and upright (positive m). The lens here is acting as a magnifying glass. (Construct a ray diagram for this example.)

A converging lens can produce a real or a virtual image, depending on the distance of the object relative to the focal length.

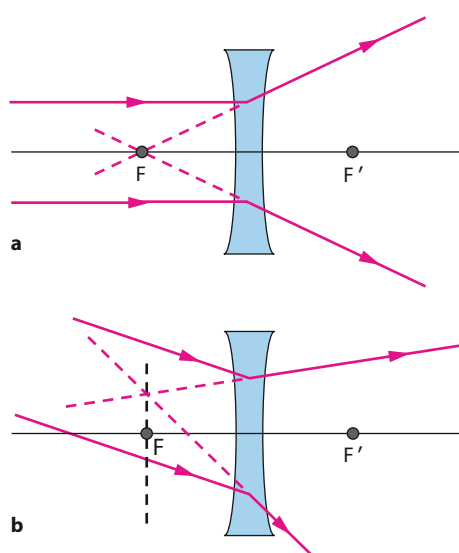


Figure C.13 **a** Rays parallel to the principal axis diverge from the lens in such a way that the extensions of the rays pass through the focal point of the lens on the same side as the incoming rays. **b** Other parallel rays at an angle to the principal axis will appear to come from a point off the principal axis at a distance from the centre of the lens equal to the focal length.

C1.3 Diverging lenses

Lenses that are thinner at the centre than at the edges are **diverging lenses**, which means that a ray of light changes its direction away from the axis of the lens (see Figure C.2b). Rays of a parallel beam of light diverge from each other after going through the lens (Figure C.13).

With small but important changes, much of the discussion for converging lenses can be repeated for diverging lenses. We need to know the behaviour of three standard rays in order to construct an image graphically.

First we need a definition of the focal point of a diverging lens. In a diverging lens, rays coming in parallel to the principal axis will, upon refraction, move away from the axis in such a way that their extensions go through a point on the principal axis called the **focal point** of the lens. The distance of the focal point from the centre of the lens is the **focal length** of the diverging lens (see Figure C.13a). This is our ‘standard ray 1’ for diverging lenses.

If the beam is not parallel to the principal axis, the extensions of the refracted rays will all go through the same point at a distance from the lens equal to the focal length (see Figure C.13b).



A ray directed at the focal point on the other side of the lens will refract parallel to the principal axis, as shown in Figure C.14. This is 'standard ray 2'.

Finally, a ray that passes through the centre of the lens is undeflected, as shown in Figure C.15. This is 'standard ray 3'.

The behaviour of all three standard rays is shown in Figure C.16.

With this knowledge we can obtain the images of objects placed in front of a diverging lens. Consider an object 8.0 cm in front of a diverging lens of focal length 6.0 cm. The height of the object is 2.0 cm. Using all three standard rays (even though only two are required), we see that an image is formed at about 3.4 cm from the lens. The image is virtual (formed by extensions of rays) and upright, and has a height of about 0.86 cm (see Figure C.17).

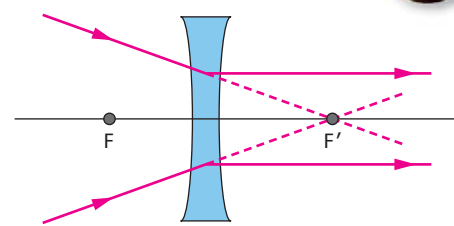


Figure C.14 A ray directed towards the focal point on the other side of the lens emerges parallel to the principal axis.

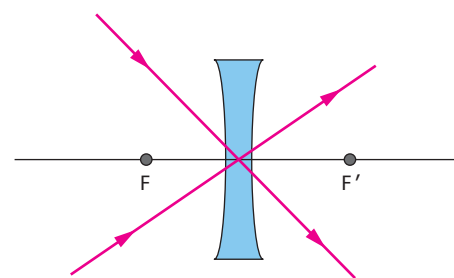


Figure C.15 A ray directed at the centre of the lens passes through undeflected.

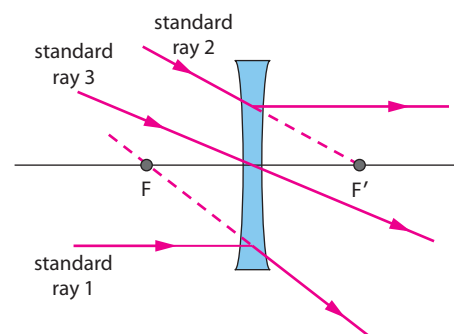


Figure C.16 Refraction of the three standard rays in a diverging lens.

Exam tip

In using the lens formula for a diverging lens, the focal length is taken to be negative.

It can be shown that the formula relating object and image distances and focal length that we used for converging lenses applies to diverging lenses as well, with the very important difference that the focal length is taken as negative. The remaining conventions are the same as for converging lenses.

As an example, consider a diverging lens of focal length 10 cm and an object placed 15 cm from the lens. Then

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = -\frac{1}{10} - \frac{1}{15} = -\frac{1}{6.0} \Rightarrow v = -6.0 \text{ cm}$$

The negative sign for v implies that the image is virtual and is formed on the same side of the lens as the object. The magnification is

$$m = -\left(\frac{-6.0}{15}\right) = +0.40, \text{ implying an upright image 40\% of the height of the object.}$$

When the object is real (i.e. $u > 0$), a diverging lens always produces a virtual image ($v < 0$). The magnification is then always positive, implying an upright image.

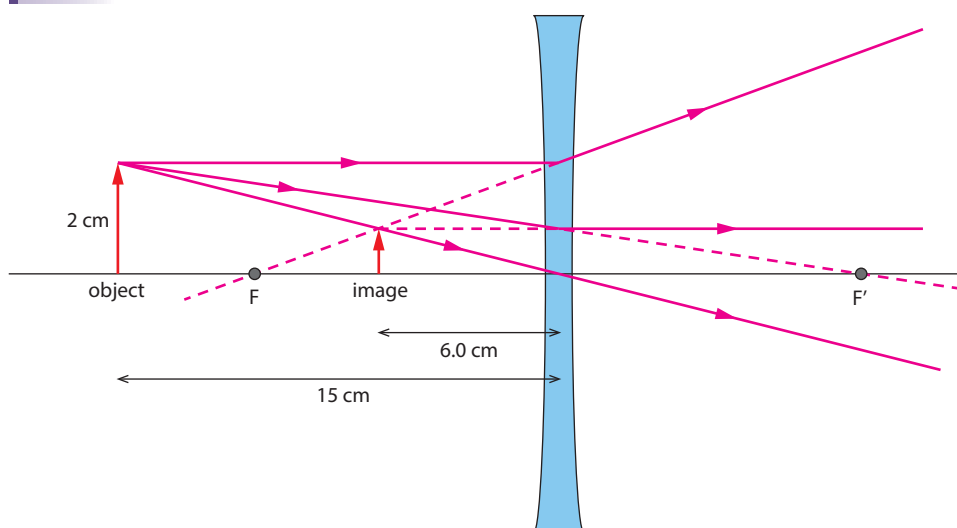
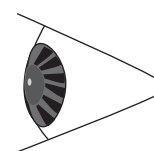


Figure C.17 Formation of an image by a diverging lens. All three standard rays are shown here.



C1.4 Lens combinations: virtual objects

Figure C.18 shows two converging lenses 12 cm apart. The left lens has a focal length of 4.0 cm and the right lens a focal length of 2.0 cm. An object 4.0 cm tall is placed 12.0 cm to the left of the left lens. What are the characteristics of the final image? The answer may be obtained by a ray diagram or algebraically. We begin with the ray diagram.

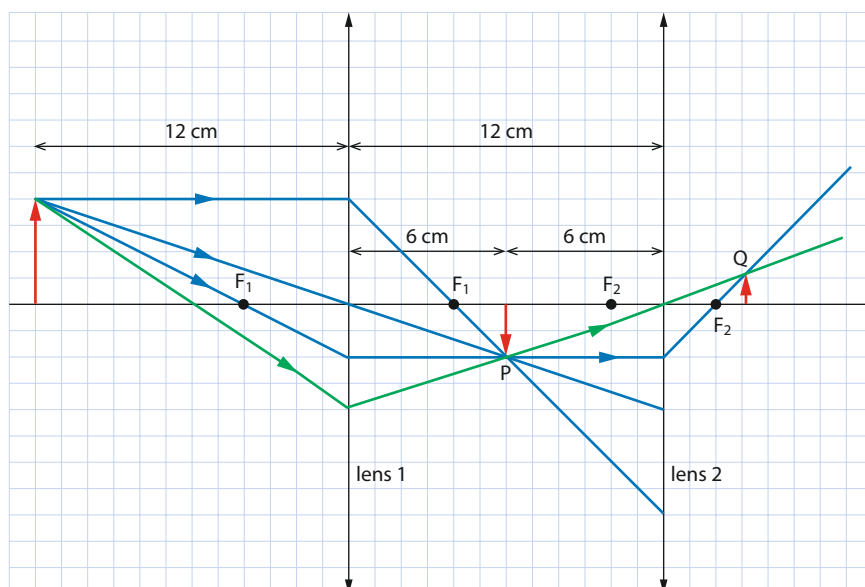


Figure C.18 Formation of a real image by a two-lens system.

We have drawn the three standard rays leaving the object (in blue). Upon refraction through the first lens the rays meet at point P and continue until they reach the second lens. The position of the image in the first lens is at P.

This image now serves as the object for the second lens. The diagram shows the three standard rays of the first lens arriving at the second lens. Of these, only one is also a standard ray for the second lens: the one parallel to the principal axis. We know that this ray, upon refraction in the second lens, will pass through the focal point of the second lens, as shown in Figure C.18. We need another ray through the second lens. We choose the one from P which passes through the centre of the second lens (the green ray), emerges undeflected and meets the blue ray at Q. This is the position of the final image. We see that this image is real, 3.0 cm to the right of the second lens, upright and with a height of 1.0 cm. (In the diagram we have extended the green line backwards to the top of the object.)

These results can also be obtained using the formula. We first find the image in the first lens: we have that $u_1 = 12$ cm and $f_1 = 4.0$ cm, so

$$\frac{1}{v_1} = \frac{1}{f_1} - \frac{1}{u_1} = \frac{1}{4.0} - \frac{1}{12} = \frac{1}{6.0} \Rightarrow v_1 = 6.0 \text{ cm}$$

The magnification of the first lens is

$$m_1 = -\frac{v_1}{u_1} = -\frac{6.0}{12} = -0.50$$

This means that the image is half the size of the object and inverted. This is what the ray diagram shows as well. This image now serves



as the object for the second lens. The distance of this object from the second lens is $12 - 6.0 = 6.0$ cm. Hence the new (and final) image is at

$$\frac{1}{v_2} = \frac{1}{f_2} - \frac{1}{u_2} = \frac{1}{2.0} - \frac{1}{6.0} = \frac{1}{3.0} \Rightarrow v_2 = 3.0 \text{ cm}$$

The magnification of the second lens is

$$m_2 = -\frac{v_2}{u_2} = -\frac{3.0}{6.0} = -0.50$$

This means that the image is inverted relative to its object. But the object is already inverted, so the final image is upright. It is 50% as high as the object, or 1.0 cm tall. Overall, the image is four times smaller than the original object. This is because the overall magnification of the two-lens system is $m = m_1 m_2 = -0.50 \times (-0.50) = +0.25$.

We now consider a slightly more involved example. We again have two converging lenses 8.0 cm apart. The focal lengths are 6.0 cm and 4.0 cm for the left and right lens, respectively. The object is again 12 cm from the left lens (Figure C.19).

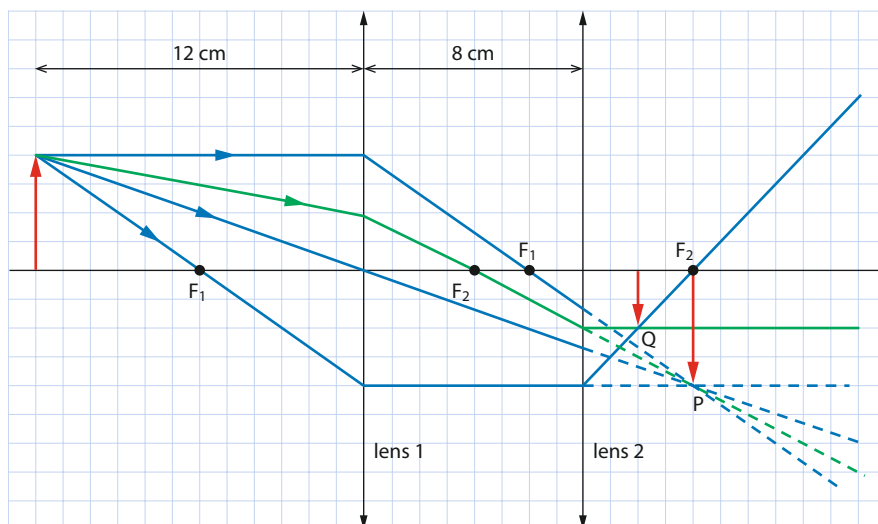


Figure C.19 Formation of a real image by a two-lens system.

We again draw the three standard rays (blue) leaving the top of the object. If the second lens were not there, these rays would meet at P and the image in the first lens would form there. Instead, the three blue rays arrive at the second lens. Of these, only one is a standard ray for the second lens: the one that is parallel to the principal axis, which will pass through the focal point to the right of the second lens. We need one more ray from the top of the original object, and choose the green ray, through focal point F_2 and point P. Since this ray passes through F_2 it will refract parallel to the principal axis. This ray intersects the blue line at Q, and this is the position of the final image. We see that the image is 2.0 cm to the right of the right lens, upright and 2.0 cm tall, and is a real image.

How do we get the same results with the formula? Applying it to the first lens, we have that $u_1 = 12$ cm and $f_1 = 6.0$ cm, so

$$\frac{1}{v_1} = \frac{1}{f_1} - \frac{1}{u_1} = \frac{1}{6.0} - \frac{1}{12} = \frac{1}{12} \Rightarrow v_1 = 12 \text{ cm}$$

Exam tip

Let h_1 be the height of the original object, h_2 the height of the image in the first lens and h_3 the height of the final image. The overall magnification is

$$m = \frac{h_3}{h_1} = \frac{h_3}{h_2} \times \frac{h_2}{h_1} = m_1 m_2,$$

which is the product of the magnifications for the individual lenses.

The image is 12 cm to the right of the first lens (or 4.0 cm to the right of the second lens). The magnification is

$$m_1 = -\frac{v_1}{u_1} = -\frac{12}{12} = -1.0$$

This means that the image is the same size as the object and inverted, consistent with our ray diagram.

Now this image serves as the object for the second lens. However, it is on the 'wrong' side of the lens! This makes this image a 'virtual object' for the second lens, and we must take its sign as negative in the formula for the second lens: $u_2 = -4.0$ cm.

$$\frac{1}{v_2} = \frac{1}{f_2} - \frac{1}{u_2} = \frac{1}{4.0} - \left(-\frac{1}{12}\right) = \frac{1}{2.0} \Rightarrow v_2 = 2.0 \text{ cm}$$

The magnification of the second lens is

$$m_2 = -\frac{v_2}{u_2} = -\frac{2.0}{-4.0} = +0.50$$

This means that the image is not inverted relative to its object. But that object is already inverted, so the final image is inverted, relative to the original object. It is half as tall as the object, or 2.0 cm tall, and thus half as tall as the original object. This is because the overall magnification of the two-lens system is $m = m_1 m_2 = -1.0 \times 0.50 = -0.50$.

Worked examples

C.4 An object lies on a table. A converging lens of focal length 6.0 cm is placed 4.0 cm above the object.

- Determine the image formed by this lens.
- A second converging lens of focal length 5.0 cm is now placed 3.0 cm above the first lens. Determine the image formed by this combination of lenses.

a With just the first lens, the image is formed at a distance found from

$$\frac{1}{v_1} + \frac{1}{u_1} = \frac{1}{f_1}$$

$$\frac{1}{v_1} = \frac{1}{6.0} - \frac{1}{4.0} = -\frac{1}{12}$$

$$v_1 = -12 \text{ cm}$$

The magnification is $m_1 = -\frac{v_1}{u_1} = -\frac{-12}{4.0} = +3.0$

The image is virtual, upright and three times taller.

b This image acts as the object for the second lens. Its distance from the second lens is $12 + 3.0 = 15$ cm. It is a real object for the second lens, so $u_2 = +15$ cm. The new image is thus formed at a distance found from

$$\frac{1}{15} = \frac{1}{v_2} - \frac{1}{5.0}$$

$$v_2 = 7.5 \text{ cm}$$

The final image is thus real. The magnification of the second lens is $m_2 = -\frac{v_2}{u_2} = -\frac{7.5}{15} = -0.50$

The overall magnification is $m = m_1 m_2 = 3.0 \times (-0.50) = -1.5$. Thus the final image is inverted and 1.5 times as tall.



C.5 An object is placed 8.0 cm to the left of a converging lens of focal length 4.0 cm. A second diverging lens of focal length 6.0 cm is placed 4.0 cm to the right of the converging lens. Determine the image of the object in the two-lens system, and verify your results with a scaled ray diagram.

The image in the converging lens is found from

$$\frac{1}{8.0} + \frac{1}{v_1} = \frac{1}{4.0} \Rightarrow v_1 = 8.0 \text{ cm}$$

Its distance from the diverging lens is therefore 4.0 cm. This image acts as a virtual object for the diverging lens. Hence $u_2 = -4.0$ cm. The final image is therefore at a distance found from

$$\frac{1}{-4.0} + \frac{1}{v_2} = \frac{1}{-6.0} \Rightarrow v_2 = 12 \text{ cm}$$

The image is thus real. The magnification of the lens system is $\left(\frac{-8.0}{8.0}\right) \times \left(\frac{-12}{-4.0}\right) = -3.0$, which implies that the final image is inverted and three times as large as the original object. Figure C.20 is a ray diagram of the problem.

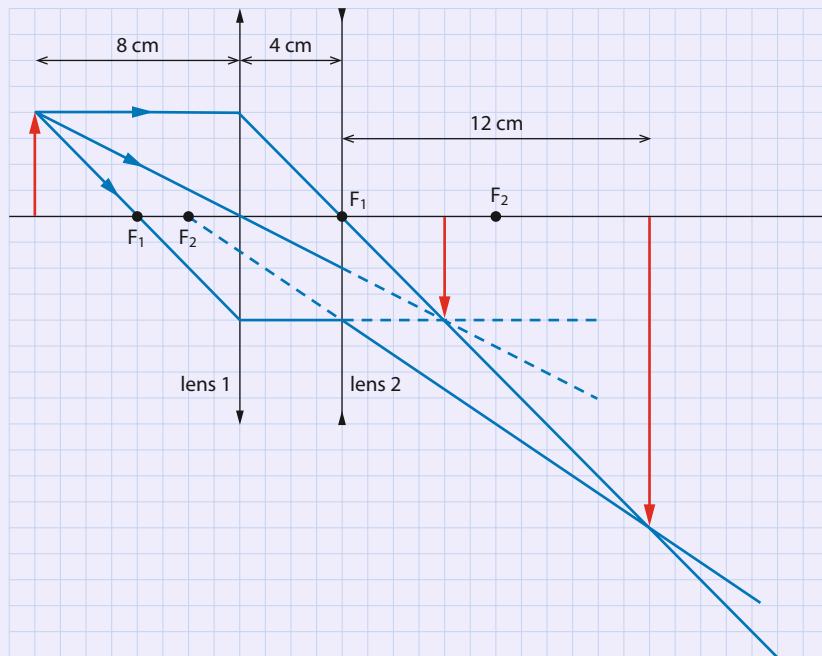


Figure C.20

C1.5 Wavefronts and lenses

Since a lens changes the direction of rays refracting through it, wavefronts (which are normal to rays) also change shape. Figure C.21 shows plane wavefronts in air approaching and entering a transparent surface that has a curved boundary.

The wavefront AC first reaches the curved boundary at point A. In one period, point A will move forward a distance equal to one wavelength in the new medium, where the speed of light is less than in air. The wavelength in the new medium will be shorter than in air. Point A will therefore move to point B. On the other hand, point C will move a longer distance in air and get to point D. Points B and D are part of the new wavefront. We see that it has to curve. The way the wavefronts curve implies that rays converge towards point F in the new medium and, from there on, the rays diverge.

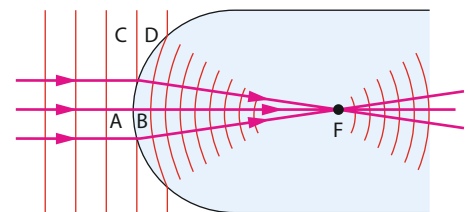


Figure C.21 Plane wavefronts curve after entering another medium with a curved boundary.

In similar fashion, we can see how wavefronts curve as they pass through converging and diverging lenses (Figure C.22).

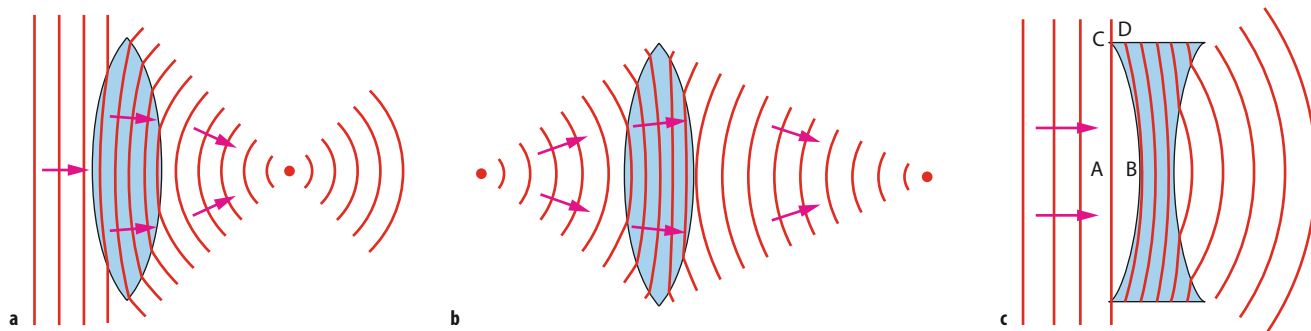


Figure C.22 **a** Plane wavefronts moving through a converging lens. **b** Spherical wavefronts moving through a converging lens. **c** Plane wavefronts moving through a diverging lens.

C1.6 Mirrors

Much of what we have learned about lenses also applies to mirrors. The big difference, of course, is that here the phenomenon is reflection of rays off a mirror surface, not refraction through a lens. We will first deal with spherical mirrors, whose surfaces are cut from a sphere.

We distinguish between concave and convex mirrors. With concave mirrors (Figure C.23a), rays parallel to the principal axis reflect through a common point on the principal axis – the focus of the mirror. With convex mirrors (Figure C.23b), rays parallel to the principal axis reflect such that their extensions go through a common point on the principal axis, behind the mirror – the focus of the convex mirror.

Figure C.24 shows how the standard rays reflect off concave and convex mirrors. Ray 1 is parallel to the principal axis. It reflects such that the ray or its extension goes through the focal point. Ray 2 goes through the focal point (or its extension does), and reflects parallel to the principal axis. Ray 3 is directed at the centre of the mirror and reflects so as to make the same angle with the principal axis as the incident ray.

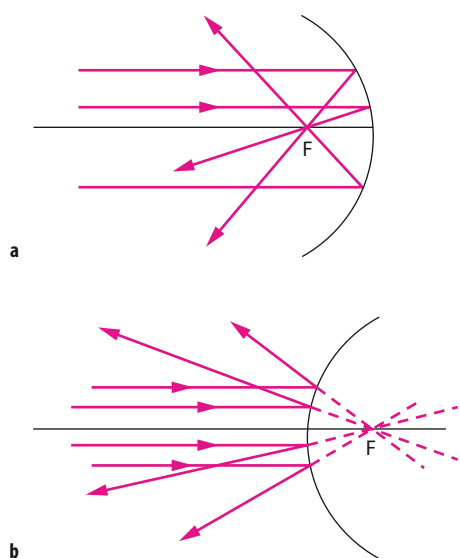


Figure C.23 **a** Concave and **b** convex spherical mirrors.

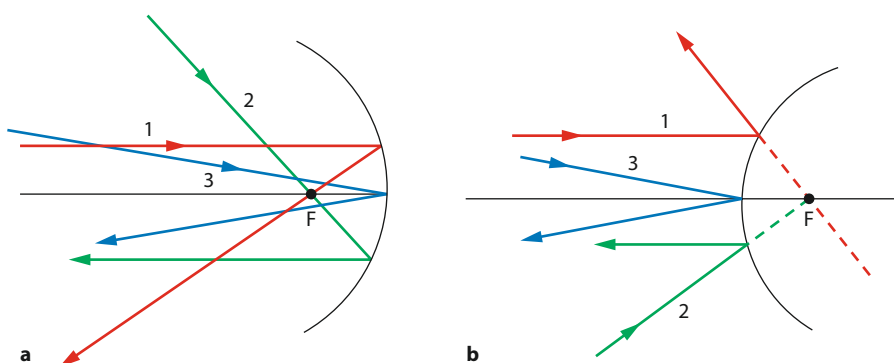


Figure C.24 The standard rays in **a** concave and **b** convex spherical mirrors.

The formula relating object and image distances to focal length that we learned for lenses also applies to mirrors, along with the same conventions. The formula for magnification is also the same. For convex mirrors (as with diverging lenses), the focal length is taken to be negative.



Worked example

C.6 Make sure you understand the formation of these images in the concave mirror in Figure C.25. Explain which case creates a real image and which a virtual image.

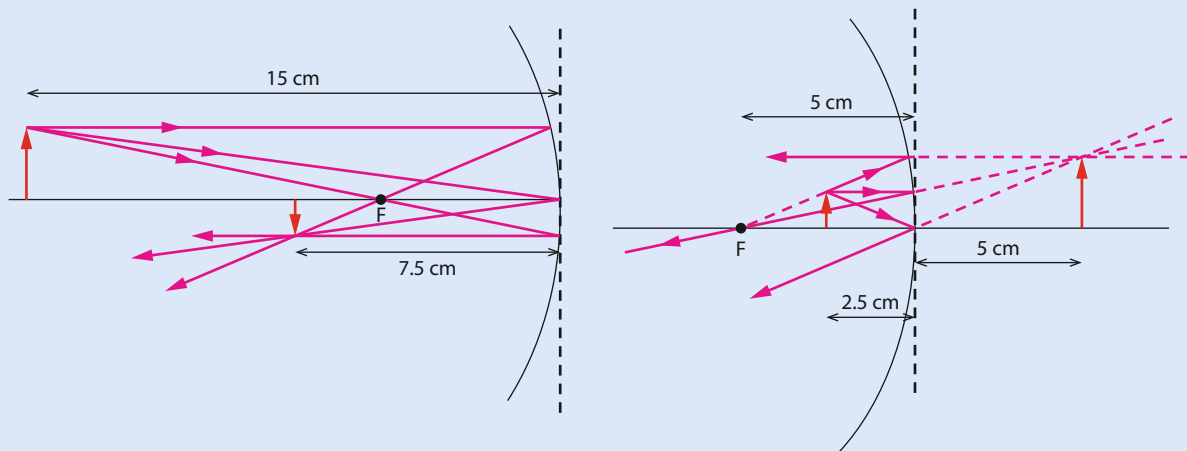


Figure C.25

The image is real in the left diagram because it is formed by real rays. The image is virtual in the second case because it is formed by ray extensions.

The statement that rays parallel to the principal axis of a spherical mirror reflect through the same point on the principal axis (the focal point) is strictly true only for rays that are very close to the principal axis; these are called **paraxial** rays. All rays parallel to the principal axis can be made to reflect through the same point using mirrors with a parabolic shape (Figure C.26).

For such mirrors, rays parallel to the axis reflect through the focal point no matter how far they are from the axis. Rays from the Sun arrive parallel to one another, and a parabolic mirror can focus them and raise the temperature at the focus, enough to heat water or even light the Olympic torch (Figure C.27).

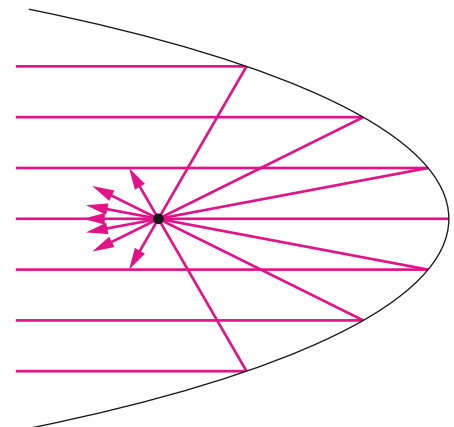


Figure C.26 With a parabolic mirror, all rays parallel to the principal axis reflect through the same point.



Figure C.27 Lighting the Olympic torch at Olympia in Greece, using a parabolic mirror to focus the rays of the Sun.

C1.7 The magnifier

The human eye can produce a clear, sharp image of any object whose distance from the eye is anything from (practically) infinite up to a point called the **near point**. Objects closer to the eye than the near point produce blurry images or force the eye to strain.

The closest point on which the human eye can focus without straining is known as the **near point** of the eye. The distance D of the near point from the eye is about 25 cm for a normal eye, but depends greatly on the age of the person involved.

The closer one gets to an object, the larger the object appears. But of course the object does not change size as you get closer! What makes it appear larger is that the angle the object subtends at the eye gets larger; this creates a larger image on the retina, which the brain interprets as a larger object.

Thus, let an observer view a small object at the near point, $D = 25$ cm from the eye, and let θ be the angle that the object subtends at the eye, as shown in Figure C.28. We know that $\tan \theta = \frac{h}{D}$, but for very small angles $\tan \theta \approx \theta$, so $\theta \approx \frac{h}{D}$.

Let us now view the object through a lens; we place the object very close to the focal point of the lens, in between the focal point and the lens. A virtual, upright, enlarged image will be formed very far from the lens (Figure C.29). The eye can then see the image comfortably and without straining; the eye is relaxed.

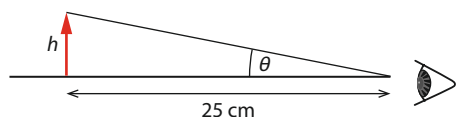


Figure C.28 The apparent size of an object depends on the angle subtended at the eye.

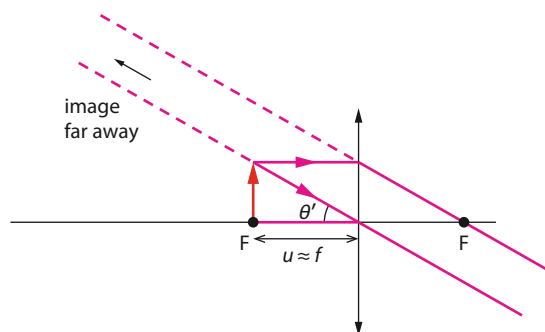


Figure C.29 A converging lens acting as a magnifier. The image is formed very far from the lens when the object is just to the right of the focal point.

Exam tip

Strictly speaking, $\tan \theta = \frac{h}{D}$.

If the angle (in radians) is very small, then $\tan \theta \approx \theta$, so

$\theta \approx \frac{h}{D}$. This is the small-angle approximation.

Exam tip

This is the formula for angular magnification when the image is formed far from the lens (practically at infinity).

We define the **angular magnification** as the ratio of the angle subtended at the eye by the image to the angle subtended by the object when is viewed by the unaided eye at the near point (Figure C.28):

$$M = \frac{\theta'}{\theta}$$

Using the small-angle approximation, $\theta = \frac{h}{D}$ and $\theta' = \frac{h}{u}$. But $u \approx f$,

$$\text{so } \theta' \approx \frac{h}{f}.$$

$$M \approx \frac{\frac{h}{f}}{\frac{h}{D}} \approx \frac{D}{f}$$



We could, however, arrange to position the object at the right place so that the image is formed at the near point (Figure C.30).

In this case the image is formed at $v = -D$ (the image is virtual, hence the minus sign), so the distance of the object from the lens is given by

$$\frac{1}{u} + \frac{1}{-D} = \frac{1}{f}$$

$$u = \frac{Df}{D+f}$$

Let θ' be the angle that the image subtends at the eye through the lens. From simple geometry we obtain

$$\theta' = \frac{h}{u} = \frac{h(D+f)}{D}$$

and $\theta \approx \frac{h}{D}$ as usual. The angular magnification M is then

$$M = \frac{\theta'}{\theta} = \frac{\frac{h(D+f)}{Df}}{\frac{h}{D}} = \frac{D+f}{f} = 1 + \frac{D}{f}$$

In both cases, the magnification can be increased by decreasing the focal length of the lens. Lens defects known as **aberrations** (see Section C1.8) limit the angular magnification to about 4.

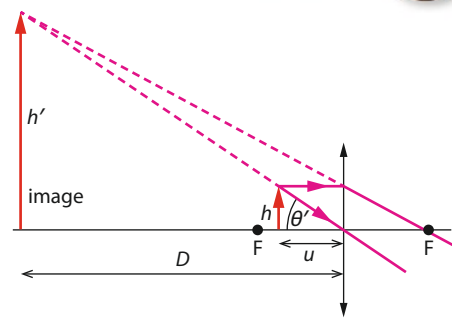


Figure C.30 A converging lens acting as a magnifier. The image is formed at the near point of the eye.

Exam tip

This is the formula for angular magnification when the image is formed at the near point.

Worked example

C.7 An object of length 4.0 mm is placed in front of a converging lens of focal length 6.0 cm. A virtual image is formed 30 cm from the lens.

- Calculate the distance of the object from the lens.
- Calculate the length of the image.
- Calculate the angular magnification of the lens.

a The object distance is found from the lens formula $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$. The image is virtual, so we must remember that $v = -30$ cm. Thus $\frac{1}{u} + \frac{1}{-30} = \frac{1}{6.0}$, so $u = 5.0$ cm.

b The linear magnification is $m = -\frac{v}{u} = -\frac{-30}{5.0} = +6.0$. The length of the object is therefore $6.0 \times 4.0 = 24$ mm.

c The angular magnification is $M = \frac{\theta'}{\theta} = \frac{\frac{h}{u}}{\frac{h}{D}} = \frac{5.0}{25} = 5.0$. The point of this is to show that you must be careful

with the formulas in the booklet. They apply to the image at infinity or at the near point. For other cases, as here, you have to work from first principles. (We can also find the image height h' from similar triangles. See Figure C.30: $\frac{h'}{30} = \frac{h}{u} \Rightarrow h' = 4.0 \times \frac{30}{5.0} = 24$ mm.)

C1.8 Lens aberrations

Lenses and mirrors do not behave exactly as described above – they suffer from **aberrations**: deviations from the simple description we have provided here. Two main types of aberration are important for lenses: **spherical** and **chromatic**.

Spherical aberration occurs because rays that enter the lens far from the principal axis have a slightly different focal length from rays entering near the axis.

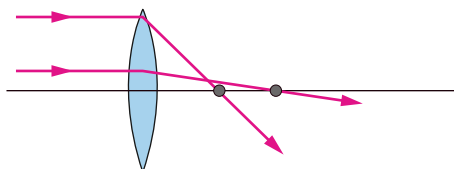


Figure C.31 Spherical aberration: rays far from the axis have a different focal point than those close to the axis, so the image is not a point. (The diagram is exaggerated.)

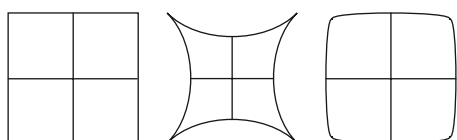


Figure C.32 An example of distortion due to spherical aberration. The grid is distorted because the magnification varies as one moves away from the principal axis.

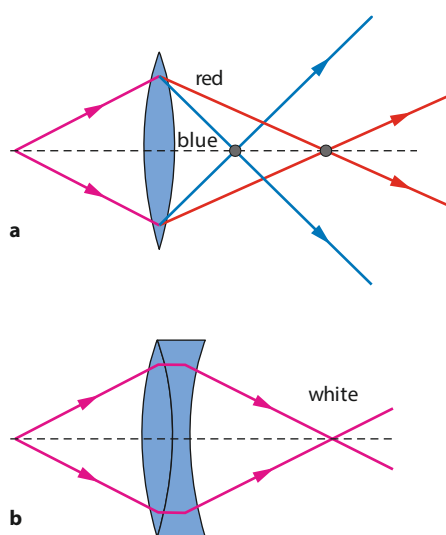


Figure C.33 a Light of different colours bends by different amounts, so different colours are focused at different places.
b An achromatic doublet consists of a pair of lenses of different indices of refraction.

In Figure C.31, rays incident on the lens far from its centre refract through a point on the principal axis that is closer to the lens than rays incident closer to the centre. This means that the image of the point is not a point but a blurred patch of light. Spherical aberration can be reduced by reducing the aperture of the lens (its diameter); this is called **stopping down**. But that means that less light goes through the lens, which results in a less bright image. And a lens with a smaller diameter would also suffer from more pronounced diffraction effects.

The fact that the focal point varies for rays that are further from the principal axis means that the magnification produced by the lens also varies. This leads to a **distortion** of the image, as shown in Figure C.32. Mirrors suffer from spherical aberration just as lenses do.

Chromatic aberration arises because the lens has different refractive indices for different wavelengths. Thus, there is a separate focal length for each wavelength (colour) of light.

This makes images appear faintly coloured – there are lines around the image in the colours of the rainbow (see Figure C.33a).

Of course, chromatic aberration disappears when monochromatic light is used. Chromatic aberration can also be reduced by combining lenses. A diverging lens with a different index of refraction placed near the first lens can eliminate the aberration for two colours and reduce it for the others (see Figure C.33b). Mirrors do not suffer from this type of aberration.

Nature of science

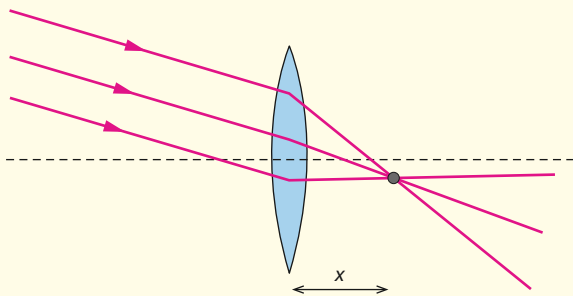
Deductive logic

What does ‘seeing an object’ mean? It means that rays from the object enter the eye, refract in the eye and finally form an image on the retina. Nerves at the back of the retina create electrical signals that are sent to the brain, and the brain reconstructs this information to create the sensation of seeing. But the rays do not necessarily have to come directly from the object and into the eye. The rays may first be reflected off a mirror or refracted through a lens. When the image formed by the mirror or lens is virtual, the brain interprets the rays as originating from a place where no actual, real object exists. Analysis of lenses and mirrors depends on this idea of the virtual image.

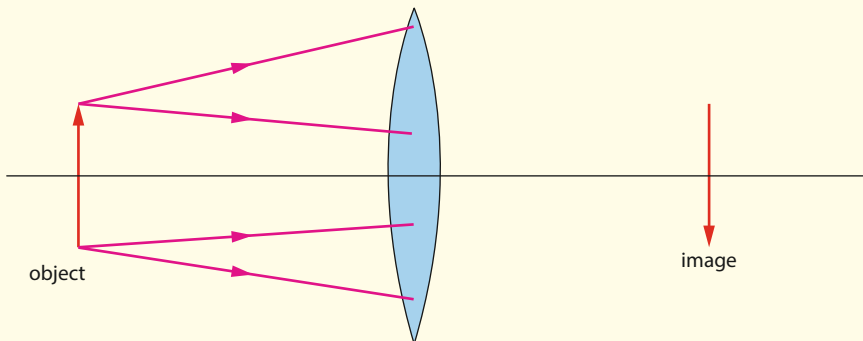


? Test yourself

- Define:
 - the **focal point of a converging lens**
 - the **focal length of a diverging lens**.
- Explain what is meant by:
 - a **real image** formed by a lens
 - a **virtual image** formed by a lens.
- Explain why a real image can be projected on a screen but a virtual image cannot.
- A plane mirror appears to reverse left and right. Does a lens do the same? Explain your answer.
- A converging lens has a focal length of 6.0 cm. Determine the distance x .

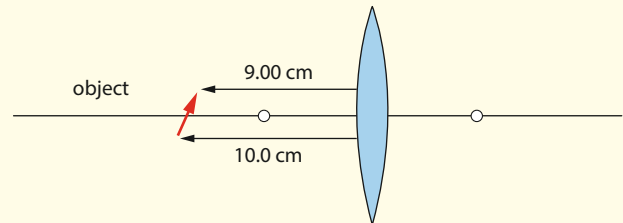


- The diagram below shows the real image of an object in a converging lens. Copy the diagram and complete the rays drawn.



- An object 2.0 cm tall is placed in front of a converging lens of focal length 10 cm. Using ray diagrams, construct the image when the object is at a distance of:
 - 20 cm
 - 10 cm
 - 5.0 cm.
 Confirm your ray diagrams by using the lens equation.
- Using a ray diagram, determine the image characteristics of an object of height 2.5 cm that is placed 8.0 cm in front of a converging lens of focal length 6.0 cm. Confirm your ray diagram by using the lens equation.

- Using a ray diagram, determine the image characteristics of an object of height 4.0 cm that is placed 6.0 cm in front of a converging lens of focal length 8.0 cm. Confirm your ray diagram by using the lens equation.
- A converging lens of focal length 4.5 cm produces a real image that is the same size as the object. Determine the distance of the object from the lens.
- Consider a converging lens of focal length 5.00 cm. An object of length 2.24 cm is placed in front of it, as shown below (not to scale), so that the middle of the object is on the principal axis. By drawing appropriate rays, determine the image in the lens. Is the angle the image makes with the principal axis the same as that for the object?

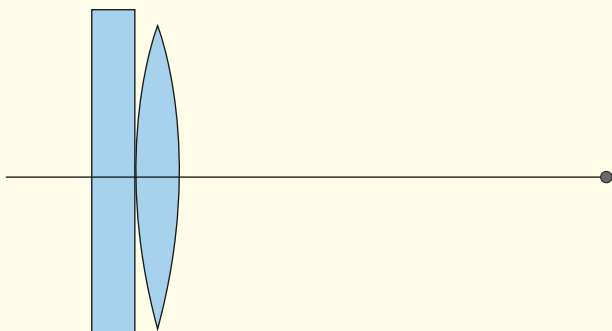


- A student finds the position of the image created by a converging lens for various positions of the object. She constructs a table of object and image distances.

$u / \text{cm} \pm 0.1 \text{ cm}$	12.0	16.0	20.0	24.0	28.0
$v / \text{cm} \pm 0.1 \text{ cm}$	60.0	27.2	19.9	17.5	16.8

- Explain how these data can be used to determine the focal length of the lens.
- Determine the focal length, including the uncertainty in its value.

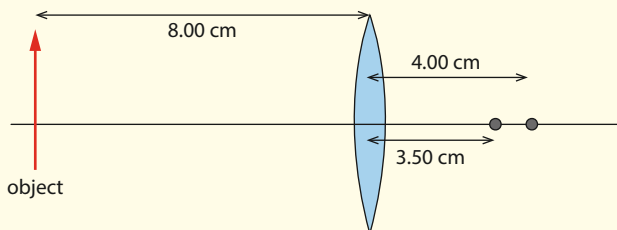
- 13 An object is placed in front of a converging lens which rests on a plane mirror, as shown below. The object is moved until the image is formed exactly at the position of the object itself. Draw rays from the object to form the image in this case. Explain how the focal length of the lens can be determined from this arrangement.



- 14 A converging lens has a focal length of 15 cm. An object is placed 20 cm from the lens.
- Determine the image (i.e. its position and whether it is real or virtual, and upright or inverted) and find the magnification.
 - Draw a ray diagram to confirm your results.
- 15 An object is 5.0 m from a screen. A converging lens of focal length 60 cm is placed between the object and the screen so that an image of the object is formed on the screen.
- Determine the distances from the screen where the lens could be placed for this to happen.
 - Determine which choice results in the larger image.
- 16 An object is placed 12 cm in front of a diverging lens of focal length 4.0 cm. Determine the properties of the image algebraically and with a ray diagram.
- 17 Two very thin lenses of focal lengths f_1 and f_2 are placed in contact. Show that the focal length of the two-lens system is given by $f = \frac{f_1 f_2}{f_1 + f_2}$.
- 18 Two converging lenses, each of focal length 10.0 cm, are 4.00 cm apart. Find the focal length of this lens combination.
- 19 An object is viewed through a system of two converging lenses, L_1 and L_2 (L_2 to the right of L_1). L_1 has a focal length of 15.0 cm and L_2 has a focal length of 2.00 cm. The distance between the lenses is 25.0 cm and the distance between the object (placed to the left of L_1) and L_1 is 40.0 cm. Determine:
- the position of the image
 - the magnification of the image
 - the orientation of the image.
- 20 An object is viewed through a system of two lenses, L_1 and L_2 (L_2 to the right of L_1). L_1 is converging and has a focal length of 35.0 cm; L_2 is diverging and has a focal length of 20.0 cm. The distance between the lenses is 25.0 cm and the distance between the object (placed to the left of L_1) and L_1 is 30.0 cm. Determine:
- the position of the image
 - the magnification of the image
 - the orientation of the image.
- 21
- An object is placed 4.0 cm in front of a concave mirror of focal length 12 cm. Determine the properties of the image.
 - Repeat part **a** when the concave mirror is replaced by a convex mirror of the same focal length.
 - In each case draw a ray diagram to show the construction of the image.
- 22 An object that is 15 mm high is placed 12 cm in front of a mirror. An upright image that is 30 mm high is formed by the mirror. Determine the focal length of the mirror and whether the mirror is concave or convex.



- 23 a Describe the two main lens aberrations and indicate how these can be corrected.
- b In an attempt to understand the distortion caused by spherical aberration, a student considers the following model. She places an object of height 4.00 cm a distance of 8.00 cm from a converging lens. One end of the object is 1.00 cm below the principal axis and the other 3.00 cm above. She assumes that rays leaving the bottom of the object will have a focal length of 4.00 cm and the rays from the top a focal length of 3.50 cm (see diagram below).
- i Under these assumptions, draw rays from the bottom and top of the object to locate the image.
- ii Draw the image again by using a 4.00 cm focal length for all rays, and compare.



- 24 An object is placed in front and to the left of a converging lens, and a real image is formed on the other side of the lens. The distance of the object from the left focal point is x and the distance of the image from the right focal point is y . Show that $xy = f^2$.

- 25 A converging lens of focal length 10.0 cm is used as a magnifying glass. An object whose size is 1.6 mm is placed at some distance from the lens so that a virtual image is formed 25 cm in front of the lens.
- a Calculate the distance between the object and the lens.
- b Suggest where the object should be placed for the image to form at infinity.
- c Find the angular size of the image at infinity.
- 26 Angular magnification, for a magnifying glass, is defined as $M = \frac{\theta'}{\theta}$.
- a By drawing suitable diagrams, show the angles that are entered into this formula.
- b A simple magnifying glass produces an image at the near point. Explain what is meant by 'near point'.
- c Show that when a simple magnifying glass produces an image at the near point, the magnification is given by $M = 1 + \frac{25}{f}$, where f is the focal length of the lens in cm.
- 27 The normal human eye can distinguish two objects 0.12 mm apart when they are placed at the near point. A simple magnifying glass of focal length 5.00 cm is used to view images at the near point. Determine how close the objects can be and still be distinguished.

C2 Imaging instrumentation

We owe much of our knowledge about the natural world to optical instruments based on mirrors and lenses. These have enabled the observation of very distant objects through telescopes and very small objects through microscopes. We have already seen how a single converging lens can produce an enlarged upright image of an object placed closer to the lens than the focal length, thus acting as a magnifying glass. The apparent size of an object depends on the size of the image that is formed on the retina. In turn, this size depends on the angle subtended by the object at the eye. This is why we bring a small object closer to the eye in order to view it – the angle subtended at the eye by the object increases.

C2.1 The optical compound microscope

A **compound microscope** (Figure C.34) consists of two converging lenses. It is used to see enlarged images of very small objects. The object (of height h) is placed at a distance from the first lens (the **objective**)

Learning objectives

- Describe and solve problems with compound microscopes.
- Describe and solve problems with astronomical refracting and reflecting telescopes.
- Outline the use of single-dish radio telescopes.
- Understand the principle of radio interferometry telescopes.
- Appreciate the advantages of satellite-borne telescopes.

which is slightly greater than the focal length f_o of the objective. A real inverted image of height h' is formed in front of the second lens (the **eyepiece**) and to the right of the focal point F_e of the eyepiece. This image serves as the object for the eyepiece lens, which, acting as a magnifier, produces an enlarged, virtual image of height h'' .

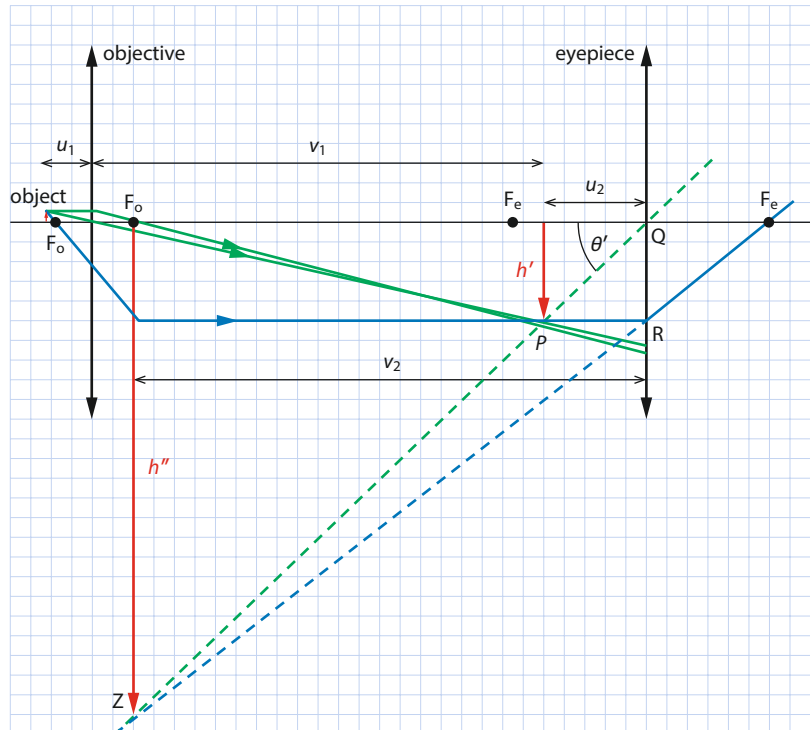


Figure C.34 A compound microscope consists of two converging lenses.

The diagram shows the intermediate image, determined by the three standard rays through the objective lens. One of the three standard rays, shown in dark blue, is also a standard ray for the eyepiece lens; from R it is refracted through the focal point of the eyepiece. To form the image, we draw the dashed green ray from P, the top of the intermediate image, to Q, the centre of the lens. Extended backwards, this intersects the extension of the dark blue ray at Z, the top of the final image.

The overall angular magnification of the microscope is defined, as usual, as the ratio of two angles: θ' , which the final image subtends at the eyepiece, to θ , which the original object would subtend when viewed from the near point distance D :

$$M = \frac{\theta'}{\theta}$$

But $\theta' \approx \frac{h''}{v_2}$ and $\theta \approx \frac{h}{D}$, so

$$M \approx \frac{h''}{h} \approx \frac{D}{f} = \frac{h'' D}{h v_2} = \frac{h'' h' D}{h' h v_2}$$

The overall angular magnification of the microscope is therefore

$$M \approx m_o \times m_e \times \frac{D}{v_2}$$

Exam tip

In a microscope we want an objective with a short focal length and an eyepiece with a long focal length.

Exam tip

General formula for angular magnification of a microscope.



where m_o and m_e are the **linear** magnifications of the objective and eyepiece lenses, respectively. The linear magnification of the objective is $m_o = -\frac{v_1}{u_1}$, where u_1 is the distance of the object and v_1 the distance of the image (in the objective) from the objective. The linear magnification of the eyepiece is $m_e = -\frac{v_2}{u_2}$. If the final image is formed at the near point (referred to as **normal adjustment**), then $v_2 = D$ and in that case $m_e = \frac{D}{f_e} + 1$ (see Section C1.7), so

$$M \approx m_o \times \left(\frac{D}{f_e} + 1 \right)$$

To understand the meaning of the angular magnification of a microscope, consider a microscope of overall angular magnification (-250) . Suppose that we are looking at an object that is $8 \mu\text{m}$ long. This object, when magnified, will **appear** to have the same size as an object of size $250 \times 8 \mu\text{m} = 2 \text{mm}$ viewed from 25cm .

Exam tip

This is the angular magnification at normal adjustment of the microscope (i.e. when the image is at the near point).

Notice that, in this case, the angular and linear magnifications of the eyepiece lens are the same.

None of these formulas is in the data booklet; you will have to derive them.

Worked examples

C.8 A compound microscope has an objective of focal length 2.0 cm and an eyepiece of focal length 6.0 cm . A small object is placed 2.4 cm from the objective. The final image is formed 25 cm from the eyepiece. Calculate **a** the distance of the image in the objective from the objective lens, and **b** the distance of this image from the eyepiece lens. **c** Determine the overall magnification of the microscope.

a We use the lens formula for the objective to get

$$\frac{1}{2.4} + \frac{1}{v_1} = \frac{1}{2.0} \Rightarrow v_1 = 12 \text{ cm}$$

b Now we do the same for the eyepiece to get

$$\frac{1}{u_2} + \frac{1}{-25} = \frac{1}{6.0} \Rightarrow u_2 = 4.8 \text{ cm}$$

c The linear magnification of the objective is $m_o = -\frac{v_1}{u_1} = -\frac{12}{2.4} = -5.0$.

The angular magnification of the eyepiece is $M_e = 1 + \frac{D}{f_e} = 1 + \frac{24}{6.0} = 5.0$. The overall magnification is therefore $-5.0 \times 5.0 = -25$.

C.9 In a compound microscope the objective has a focal length of 1.0 cm and the eyepiece a focal length of 4.0 cm . A small object is placed 1.2 cm from the objective. The final image is formed 30 cm from the eyepiece. Calculate the magnification of the microscope.

Applying the formula $M_e \approx m_o \times m_e \times \frac{D}{v_2}$, we find

$$\frac{1}{1.2} + \frac{1}{v_1} = \frac{1}{1.0} \Rightarrow v_1 = 6.0 \text{ cm}$$

Thus, $m_o = -\frac{6.0}{1.2} = -5.0$.

$$\frac{1}{u_2} + \frac{1}{-30} = \frac{1}{5.0} \Rightarrow u_2 = 4.29 \text{ cm}$$

Thus, $m_e = -\frac{-30}{4.29} = 7.0$, so $M \approx -5.0 \times 7.0 \times \frac{25}{30} \approx 29$.

Exam tip

We would like to have a large angle α because the larger this angle the more light is collected by the microscope. So to reduce d_{\min} we have to put a medium of high refractive index between the object and the lens. Such microscopes are called oil immersion microscopes.

(You do not need to know this formula for the exam.)

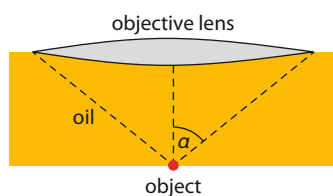


Figure C.35 An oil immersion microscope has a higher resolution than one without the oil.

Exam tip

In a telescope we want to have an objective with a long focal length and an eyepiece with a short focal length.

Exam tip

Formula for angular magnification of a refracting telescope with image at infinity (normal adjustment).

C2.2 Resolution of a compound microscope

Diffraction means that a point source will not have a point image in a lens; the image will be a disc of spread-out light. Thus if two point sources are very close to each other, their images will overlap and so may not be seen as distinct. This limits the resolution of the microscope. It can be shown that the smallest distance that can be resolved in a microscope is

$$d_{\min} = \frac{0.61\lambda}{n \sin \alpha}$$

where λ is the wavelength of the light and α is the angle shown in Figure C.35. There is a very small quantity of oil of refractive index n between the objective lens and the object. Since $n > 1$, this makes d_{\min} smaller than what it would be without the oil – that is, it increases the microscope's resolution.

C2.3 The refracting telescope

The function of a telescope is to allow the observation of large objects that are very distant and so appear very small. A **star** is enormous but looks small because it is far away. The telescope increases the angle subtended by the star relative to the angle subtended at the unaided eye. The telescope does not provide **linear** magnification of the star, since in that case the image would be many orders of magnitude larger than the Earth!

A refracting astronomical telescope (Figure C.36) consists of two converging lenses. Since the object observed is very far away, the image produced by the first lens (the objective) is at the focal plane of the objective. It is this image that is then magnified by the eyepiece, just as by a magnifying glass. The second lens (the eyepiece) forms a virtual, inverted image of the object. The final image is produced at infinity, so the distance between the two lenses is the sum of their focal lengths. Under these conditions the telescope is said to be in **normal adjustment**.

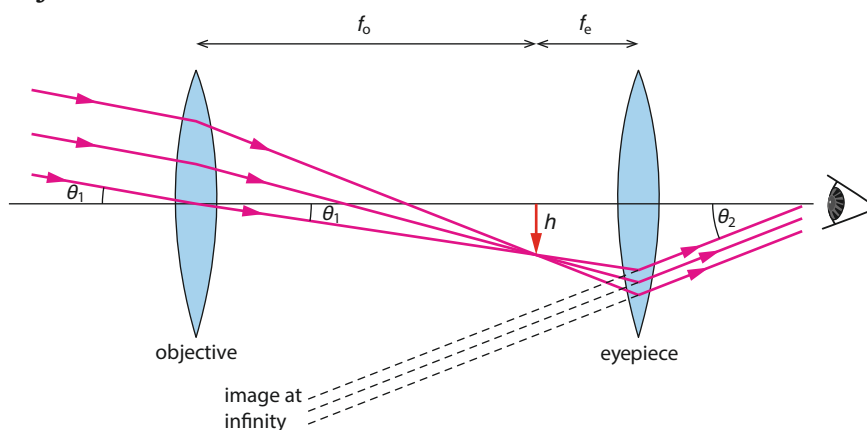


Figure C.36 A refracting astronomical telescope. The final image is inverted.

The **angular magnification** of the telescope is defined as the ratio of the angle subtended by the object as seen through the telescope to the angle subtended by it at the unaided eye. Thus:

$$M = \frac{\theta_2}{\theta_1} = \frac{h/f_e}{h/f_o} = \frac{f_o}{f_e}$$



The position of the eyepiece can be adjusted to provide clear images of objects other than very distant ones. The objective lens should be as large as possible in order to allow as much light as possible into the telescope. Because it is difficult to make very large lenses, telescopes have been designed to use mirrors rather than lenses.

Worked examples

C.10 A refracting telescope has a magnification of 70.0 and the two lenses are 60.0 cm apart at normal adjustment. Determine the focal lengths of the lenses.

The angular magnification is

$$M = \frac{f_o}{f_e} = 70 \Rightarrow f_o = 70f_e$$

so

$$f_o + f_e = 70f_e + f_e = 71f_e = 60 \text{ cm}$$

$$f_e = 0.845 \text{ cm}$$

$$f_o = 59.2 \text{ cm}$$

C.11 An astronomical telescope is used to view an object 20 m from the objective. The final real image is formed 30 cm from the eyepiece lens. The focal length of the objective is 4.0 m and that of the eyepiece is 0.80 m. Determine the overall linear magnification of the telescope.

The image in the objective is formed at a distance found from $\frac{1}{20} + \frac{1}{v_1} = \frac{1}{4.0} \Rightarrow v_1 = 5.0 \text{ m}$. Hence the linear magnification of the objective is $m_o = -\frac{5.0}{20} = -0.25$.

The object for the eyepiece is at a distance found from $\frac{1}{u_2} + \frac{1}{0.30} = \frac{1}{0.80} \Rightarrow u_2 = -0.48 \text{ m}$. Hence the linear magnification of the eyepiece is

$$m_e = -\frac{0.30}{-0.48} = 0.625.$$

The overall magnification is therefore $-0.25 \times 0.625 = -0.16$.

C2.4 Reflecting telescopes

Reflecting telescopes use mirrors rather than the lenses of **refracting telescopes**. This creates a number of advantages, including:

- To see distant faint objects requires large lenses (to collect more light). But large lenses are hard to make (the glass must be homogeneous and free of air bubbles); they can also only be supported along their rim, and large lenses may collapse under their own weight. By contrast, large mirrors can be supported along the rim and at the back.
- Mirrors do not suffer from chromatic aberration.
- Only one side has to be ground, as opposed to two for lenses.

For these reasons, the largest telescopes are reflecting.

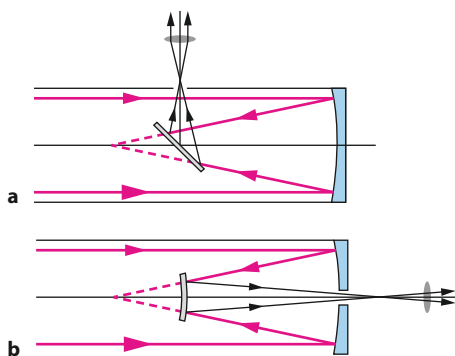


Figure C.37 Reflecting telescopes: **a** Newtonian; **b** Cassegrain.

Figure C.37 shows two types of refracting telescope. In the first, known as **Newtonian**, light from a distant object is reflected from a parabolic mirror onto a smaller plane mirror at 45° to the axis of the telescope. The reflected light is collected by a converging lens which creates a parallel beam to the observer's eye. In the second type, known as the **Cassegrain** type, light is reflected from a parabolic mirror onto a much smaller convex mirror. Light reflecting off this mirror is collected by a converging lens that produces a parallel beam to the observer's eye.

C2.5 Single-dish radio telescopes

A **radio telescope** receives and detects electromagnetic waves in the radiofrequency region. Stars, galaxies and other objects are known to radiate in this region, so studying these emissions gives valuable information about the 'invisible' side of these objects. Recall from Topic 9 that diffraction places limits on resolution, that is, on the ability of an instrument to see two nearby objects as distinct. An instrument whose diameter is b and operates at a wavelength λ can resolve two objects whose angular separation (in radians) is θ_A if

$$\theta_A \geq 1.22 \frac{\lambda}{b}$$

Since radio wavelengths are large, the diameter of the radio telescope has to be large as well, in order to achieve reasonable resolution. The Arecibo radio telescope (Figure C.38) has a diameter of 300 m and operates at a wavelength of 21 cm. This means that it can resolve objects whose angular separation is no less than

$$\theta_A \approx \frac{1.22 \times 0.21}{300} \approx 8.5 \times 10^{-4} \text{ rad}$$

By contrast, an optical telescope such as the Hubble Space Telescope (HST) has a diameter of 2.4 m and operates at an average optical wavelength of 500 nm, so

$$\theta_A \approx \frac{1.22 \times 500 \times 10^{-9}}{2.4} \approx 2.5 \times 10^{-7} \text{ rad}$$

But large telescopes are very heavy steel structures, and difficult to steer. The Arecibo telescope is actually built into a valley and cannot be steered at all: it points to different parts of the sky only because the Earth rotates.

Radio telescopes have a parabolic shape. Parallel rays will therefore collect at the focus of the mirror, where a detector is placed.



Figure C.38 The Arecibo radio telescope in Puerto Rico.

C2.6 Radio interferometry telescopes

The low resolution of single-dish radio telescopes can be overcome by a technique known as **radio interferometry**. By using a very large array of radio telescopes very far apart and appropriately combining the signals from the individual dishes, one can achieve the same resolution as a single dish with a diameter equal to the length of the array. The Very Large Array (VLA) interferometer has 27 single dishes extending over 35 km (Figure C.39). It operates at a wavelength of 6 cm so its resolution is

$$\theta_A \approx \frac{1.22 \times 0.06}{35 \times 10^3} \approx 2.1 \times 10^{-6} \text{ rad}$$

This is only about 10 times lower than the HST.

C2.7 Satellite-borne telescopes

Earthbound telescopes are limited for a number of reasons, including:

- Light pollution (excess light in the atmosphere). This can be partly overcome by locating telescopes in remote areas, far from large cities.
- Atmospheric turbulence (mainly due to convection currents and temperature differences). This makes air move unpredictably, making the positions of stars appear to vary. It can be partly overcome by locating telescopes on high mountains, where the atmosphere tends to be more stable.
- Absorption of various wavelengths by the atmosphere. This makes observation at these wavelengths impossible. This is especially true for X-ray and ultraviolet wavelengths, which are almost completely absorbed by the atmosphere.

These problems do not exist for satellite-based telescopes in orbit around the Earth. The Hubble Space Telescope, shown in Figure C.40, a joint project of the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA), has truly revolutionised astronomy, and cosmology in particular, with its wealth of detailed images that have led to new discoveries and new areas of research.



Figure C.40 The Hubble Space Telescope in orbit.



Figure C.39 The Very Large Array in New Mexico, USA.

The spectacular image in Figure C.41 shows the supernova Cassiopeia A, a dying star. This image is a combination of images at different wavelengths: optical from the HST (yellow), infrared from the Spitzer space observatory (red) and X-rays from the Chandra X-ray observatory (blue).

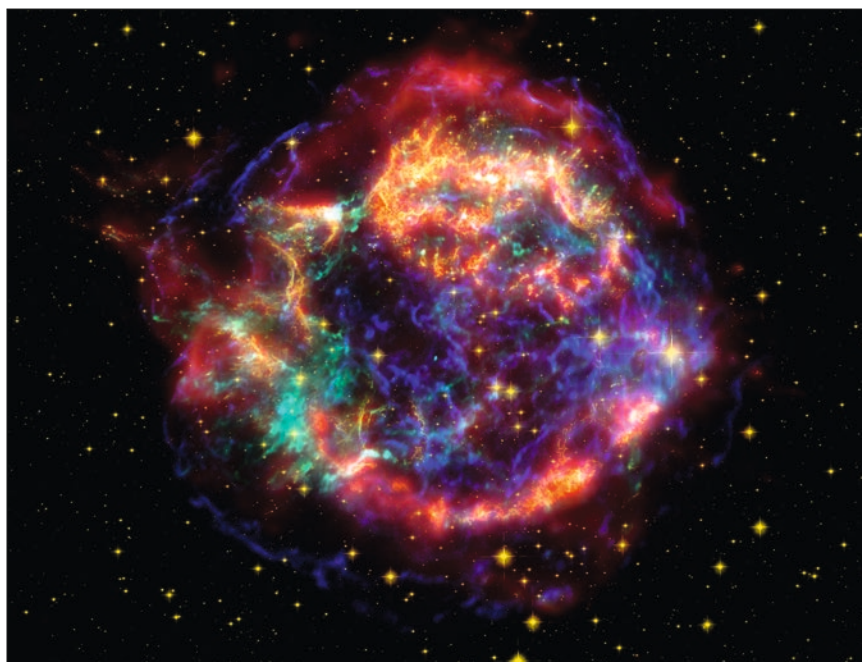


Figure C.41 A dying star, Cassiopeia A, after a supernova explosion more than 300 years ago.

Nature of science

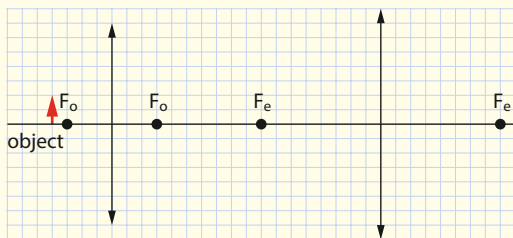
Improved instrumentation

The photograph in Figure C.41 is an excellent example of the advances in imaging made by combining data from telescopes operating at different wavelengths. Observations that until recently were only made with optical telescopes on the Earth are now complemented by images from telescopes in space operating in the radio, infrared, ultraviolet, X-ray and gamma-ray regions of the electromagnetic spectrum. Placing telescopes away from the Earth's surface avoids the distorting effects of the Earth's atmosphere, and corrective optics enhances images obtained from observatories on the Earth. These developments have vastly increased our knowledge of the universe and have made possible discoveries and the development of theories about the structure of the universe that have exceeded even the most optimistic expectations. In exactly the same way, optical, electron and tunnelling microscopes have advanced our knowledge of the biological world, leading to spectacular advances in medicine and the treatment of disease.

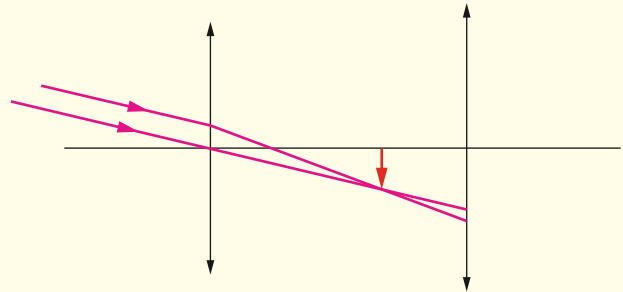


? Test yourself

- 28 The objective of a microscope has a focal length of 0.80 cm and the eyepiece has a focal length of 4.0 cm. An object is placed 1.50 cm from the objective. The final image is formed at the near point of the eye (25 cm).
- Calculate the distance of the image from the objective.
 - Calculate the distance from the eyepiece lens of the image in **a**.
 - Calculate the angular magnification of the microscope.
- 29 In a compound microscope the objective focal length is 20 mm and the eyepiece focal length is 80 mm. An object is placed 25 mm from the objective. The final virtual image is formed 35 cm from the eyepiece.
- Calculate the distance of the image from the objective.
 - Calculate the distance from the eyepiece lens of the image in **a**.
 - Calculate the angular magnification of the microscope.
- 30 The diagram below illustrates a compound microscope. Copy the diagram and draw rays in order to construct the final image.



- 31 A compound microscope forms the final image at a distance of 25 cm from the eyepiece. The eye is very close to the eyepiece. The objective focal length is 24 mm and the object is placed 30 mm from the objective. The angular magnification of the microscope is 30. Determine the focal length of the eyepiece.
- 32 The diagram below shows rays from a distant object arriving at a refracting telescope. Copy the diagram and complete the rays to show the formation of the final image at infinity.



- 33 An astronomical telescope is in normal adjustment.
- State what is meant by this statement.
 - The angular magnification of the telescope is 14 and the focal length of the objective is 2.0 m. Calculate the focal length of the eyepiece.
- 34 The Moon is at a distance of 3.8×10^8 m from the Earth and its diameter is 3.5×10^6 m.
- Show that the angle subtended by the diameter of the Moon at the eye of an observer on the Earth is 0.0092 rad.
 - A telescope objective lens has a focal length of 3.6 m and an eyepiece focal length of 0.12 m. Calculate the angular diameter of the image of the Moon formed by this telescope.
- 35 A telescope consists of an objective, which is a converging lens of focal length 80.0 cm, and the eyepiece of has a focal length 20.0 cm. The object is very far from the objective (effectively an infinite distance away) and the image is formed at infinity.
- Calculate the angular magnification of this telescope.
 - The telescope is used to view a building of height 65.0 m a distance of 2.50 km away. Calculate the angular size of the final image.
- 36 A refracting telescope has an eyepiece of focal length 3.0 cm and an objective of focal length 67.0 cm.
- Calculate the magnification of the telescope.
 - State the length of the telescope. (Assume that the final image is produced at infinity.)

- 37 A refracting telescope has a distance between the objective and the eyepiece of 60 cm. The focal length of the eyepiece is 3.0 cm. The eyepiece has to be moved 1.5 cm further from the objective to provide a clear image of an object some finite distance away. Estimate this distance. (Assume that the final image is produced at infinity.)
- 38 a State what is meant by radio interferometry.
 b Estimate the resolution in radians of an array of radio telescopes extending over 25 km. The telescopes operate at a wavelength of 21 cm.
 c Estimate the smallest separation that can be resolved by this array, in a galaxy that is 2×10^{22} m from the Earth.
- 39 Suggest why telescopes other than optical ones are in use.
- 40 Suggest why parabolic mirrors are used in telescopes.
- 41 State **two** advantages and **two** disadvantages of satellite-based telescopes.

Learning objectives

- Describe optical fibres and solve problems dealing with them.
- Describe the differences between step-index and graded-index fibres.
- Describe the difference between waveguide and material dispersion.
- Solve problems involving attenuation and the decibel scale.

C3 Fibre optics

This section introduces one very important channel of communication, the **optical fibre**. We discuss the optics of the optical fibre and the concept of the critical angle. We introduce two types of optical fibres, multimode and monomode fibres, and discuss these in the context of dispersion and attenuation.

C3.1 Total internal reflection and optical fibres

The phenomenon of **total internal reflection** was discussed in Topic 4. Here we summarise the main results. Figure C.42 shows a ray of light entering a medium of low refractive index from a medium of higher refractive index. The angle of refraction is greater than the angle of incidence. As the angle of incidence increase, the angle of refraction will eventually become 90° . The angle of incidence at which this happens is called the **critical angle**.

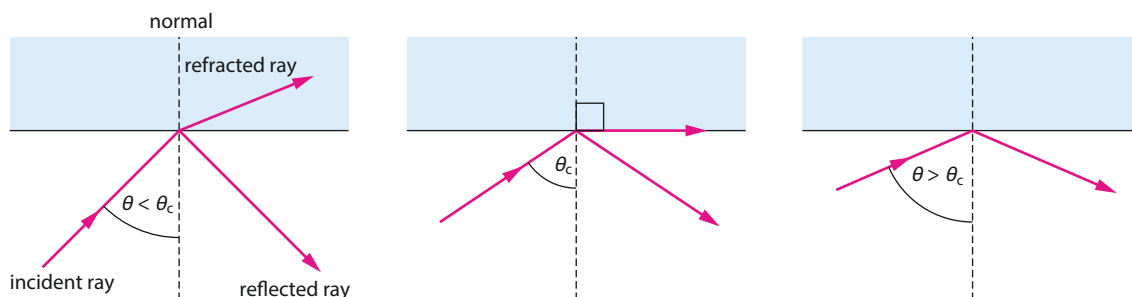


Figure C.42 A ray of light incident on a boundary partly reflects and partly refracts. The angle of refraction is larger than the angle of incidence.

The critical angle is the angle of incidence for which the angle of refraction is 90° .



The critical angle θ_c can be found from Snell's law:

$$n_1 \sin \theta_c = n_2 \sin 90^\circ$$

$$\sin \theta_c = \frac{n_2}{n_1}$$

$$\theta_c = \arcsin \frac{n_2}{n_1}$$

For an angle of incidence greater than the critical angle, no refraction takes place. The ray is simply reflected back into the medium from which it came. This is called **total internal reflection**.

One important application of total internal reflection is a device known as an **optical fibre**. This consists of a very thin glass core surrounded by a material of slightly lower refractive index (the **cladding**). Such a thin fibre can easily be bent without breaking, and a ray of light can be sent down the length of the fibre's core. For most angles of incidence, total internal reflection occurs (Figure C.43), so the light ray stays within the core and never enters the cladding.

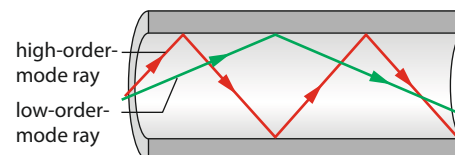


Figure C.43 A ray of light follows the shape of an optical fibre by repeated internal reflections.

Worked example

C.12 The refractive index of the core of an optical fibre is 1.50 and that of the cladding is 1.40. Calculate the critical angle at the core–cladding boundary.

From Snell's law, we have

$$1.50 \sin \theta_c = 1.40 \sin 90^\circ$$

$$\sin \theta_c = \frac{1.40}{1.50} = 0.9333$$

$$\theta_c = \arcsin 0.9333 = 69.0^\circ$$

C3.2 Dispersion

Because the refractive index of a medium depends on the wavelength of the light travelling through it, light of different wavelengths will travel through the glass core of an optical fibre at different speeds. This is known as **material dispersion**. Therefore, a set of light rays of different wavelengths will reach the end of a fibre at different times, even if they follow the same path.

Consider a pulse of light created by turning on, say, a light-emitting diode (LED) for a short interval of time. The power of the signal as a function of time as it enters the fibre is represented on the left-hand side of Figure C.44. The area under the pulse is the energy carried by the pulse. In the output pulse, on the right-hand side of Figure C.44, the area is somewhat smaller because some energy has been lost during transmission. Because of the different travel times, the pulse has become wider.

Rays of light entering an optical fibre will, in general, follow different paths. Rays that undergo very many internal reflections over a given distance are said to follow **high-order-mode** paths, while those suffering fewer reflections follow **low-order-mode** paths (Figure C.45).

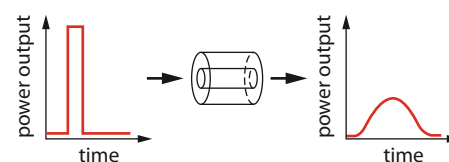


Figure C.44 The effect of material dispersion.

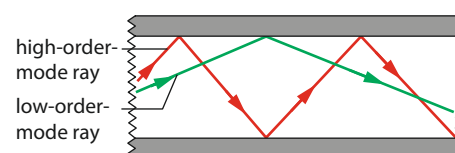


Figure C.45 Low-order- and high-order-mode rays in an optical fibre.

Consider a set of rays that have the same wavelength but follow different paths (i.e. they have different-order modes). Those rays travelling along low-order paths are more 'straight', travel a shorter distance, and so will reach the end faster than higher-order rays. This leads to what is called **waveguide dispersion**. The effect on the input signal of Figure C.44 is the same.

In practice, a set of rays will have different wavelengths and will follow different paths, so they will be subject to both material and waveguide dispersion. This is the case in **multimode** fibres (Figure C.46a and b). Multimode fibres have a core diameter of roughly $100\mu\text{m}$.

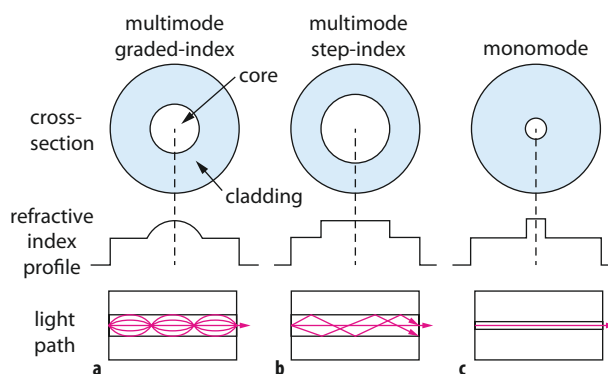


Figure C.46 a A multimode graded-index optical fibre. b A multimode step-index optical fibre. c A monomode optical fibre.

Of special interest are **monomode** fibres (Figure C.46c), in which all light propagates (approximately) along the same path. The diameter of the core of a monomode fibre is very small, about $10\mu\text{m}$, only a few times larger than the wavelength of the light entering it. The thickness of the cladding is correspondingly much larger, in order to make connecting one fibre to another easier. The propagation of light in such a fibre is not governed by the conventional laws of optics that we are using in this section. The full electromagnetic theory of light must be used, which results in the conclusion that light follows, essentially, just one path down the fibre, eliminating waveguide dispersion. Monomode fibres are now used for long-distance transmission of both analogue and digital signals.

Figure C.46b illustrates the meaning of the term **step-index fibre**. This means that the refractive index of the core is constant, and so is that of the cladding, but at a slightly lower value. The refractive index thus shows a 'step' (down) as we move from the core to the cladding. This type of fibre is to be contrasted with a **graded-index fibre**, in which the refractive index decreases smoothly from the centre of the core (where it reaches a maximum) to the outer edge of the core. The refractive index in the cladding is constant.

Exam tip

Graded index fibres help reduce waveguide dispersion: ordinarily, rays that move far from the central axis would take longer to arrive leading to dispersion; but in a graded-index fibre the speed of light away from the axis is also greater and so the longer path is covered at higher speed. The net effect is an almost constant arrival time independent of path.



Worked example

C.13 The length of an optical fibre is 5.0 km. The refractive index of the core of the optical fibre is 1.50 and the critical angle of the core–cladding boundary is 75° . Calculate the time taken for a ray of light to travel down the length of the fibre:

- along a straight line parallel to the axis of the fibre
- suffering the maximum number of internal reflections in the fibre.

The speed of light in the core of the fibre is determined by the refractive index:

$$c = \frac{3.00 \times 10^8}{1.50} = 2.00 \times 10^8 \text{ m s}^{-1}$$

a The distance travelled by the light in this case is 5.0 km, so the time taken is

$$t = \frac{5.0 \times 10^3}{2.00 \times 10^8} = 25 \mu\text{s}$$

b The ray must travel as shown in Figure C.47, with the angle θ_c being infinitesimally larger than 75° .

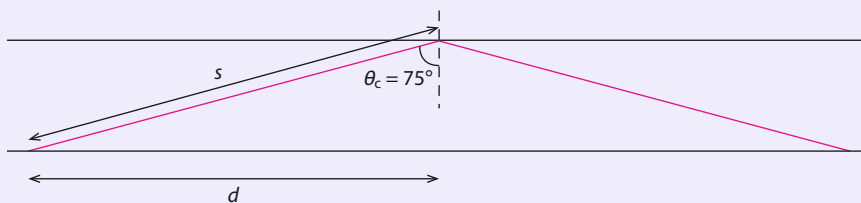


Figure C.47

Then $s = \frac{d}{\sin \theta}$. The total distance travelled by the ray is then $s = \frac{5.00}{\sin 75^\circ} = 5.18 \text{ km}$, and the time taken is

$$t = \frac{5.18 \times 10^3}{2.00 \times 10^8} = 26 \mu\text{s}.$$

C3.3 Attenuation

Any signal travelling through a medium will suffer a loss of power. This is called **attenuation**. It may be necessary to amplify the signal for further transmission. In the case of optical fibres, attenuation is mainly due to the scattering of light by glass molecules and impurities. The massive introduction of optical fibres into communications has been made possible by advances in the manufacture of very pure glass. For example, the glass in the window of a house appears to let light through without much absorption of energy, but a window pane is less than 1 cm thick. Glass of the same quality as that in ordinary windows and with a thickness of a few kilometres would not transmit any light at all.

Attenuation in an optical fibre is caused by the scattering of light and the impurities in the glass core. The amount of attenuation depends on the wavelength of light being transmitted.

To quantify attenuation, we use a logarithmic scale or **decibel scale**. We define the **power loss in decibels** (dB) as

$$\text{power loss (in dB)} = 10 \log \frac{P_{\text{final}}}{P_{\text{initial}}}$$

This is a negative quantity; P is the power of the signal.

Thus a power loss of 16 dB means that an initial power of, say, 8.0 mW has been reduced to a value given by

$$-16 = 10 \log \frac{P_{\text{final}}}{P_{\text{initial}}}$$

$$-1.6 = \log \frac{P_{\text{final}}}{8.0}$$

$$\frac{P_{\text{final}}}{8.0} = 10^{-1.6}$$

$$P_{\text{final}} = 8.0 \times 10^{-1.6} \approx 0.20 \text{ mW}$$

This idea can also be applied to signals that are amplified, as the next example shows.

Worked example

C.14 An amplifier amplifies an incoming signal of power 0.34 mW to 2.2 mW. Calculate the power **gain** of the amplifier in decibels.

The amplifier is shown schematically in Figure C.48.

For this amplifier, we have

$$\text{gain} = 10 \log \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \log \frac{2.2}{0.34} = 10 \times 0.81 = 8.1 \text{ dB}$$

It is worth remembering that an increase in power by a factor of 2 results in a power gain of approximately 3 dB:

$$\text{gain} = 10 \log \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \log 2 = 3.01 \approx 3 \text{ dB}$$

Similarly, a decrease in power by a factor of 2 implies a 3 dB power loss.

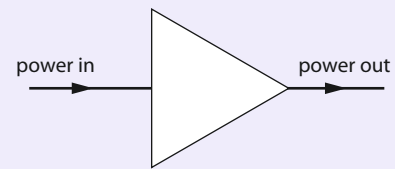


Figure C.48

Also useful is the concept of **specific attenuation**, the power loss in decibels per unit length travelled: $\text{specific attenuation} = \frac{10 \log P_{\text{out}}/P_{\text{in}}}{L}$. This is measured in decibels per kilometre (dB km^{-1}).

Worked examples

C.15 A signal of power 12 mW is input into a cable of specific attenuation 4.0 dB km^{-1} . Calculate the power of the signal after it has travelled 6.0 km in the cable.

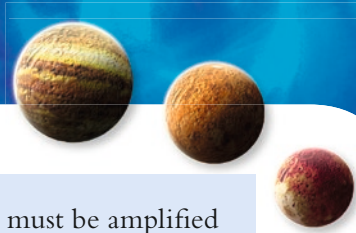
The loss is $-4.0 \times 6.0 = -24 \text{ dB}$. Then

$$-24 = 10 \log \frac{P_{\text{out}}}{P_{\text{in}}}$$

$$-2.4 = \log \frac{P_{\text{out}}}{P_{\text{in}}}$$

$$\frac{P_{\text{out}}}{P_{\text{in}}} = 10^{-2.4} \approx 3.98 \times 10^{-3}$$

$$P_{\text{out}} = 3.98 \times 10^{-3} P_{\text{in}} = 3.98 \times 10^{-3} \times 12 = 0.048 \text{ mW}$$



C.16 A signal travels along a monomode fibre of specific attenuation 3.0 dB km^{-1} . The signal must be amplified when the power has decreased by a factor of 10^{18} . Calculate the distance at which the signal must be amplified.

We know that $\frac{P_{\text{out}}}{P_{\text{in}}} = 10^{-18}$. Therefore the loss is $10 \log \frac{P_{\text{out}}}{P_{\text{in}}}$. Hence we need amplification after $\frac{180 \text{ dB}}{3.0 \text{ dB km}^{-1}} = 60 \text{ km}$.

The specific attenuation (i.e. the power loss in dB per unit length) actually depends on the wavelength of the radiation travelling along the optical fibre. Figure C.49 is a plot of specific attenuation as a function of wavelength. The graph shows minima at wavelengths of 1310 nm and 1550 nm, which implies that these are desirable wavelengths for optimal transmission. These are infrared wavelengths.

C3.4 Advantages of optical fibres

In the early days of telephone communications, signals were carried by twisted pairs of wires (Figure C.50a).

As the name suggests, pairs of copper wires were twisted around each other. This reduces noise from induced currents caused by magnetic fields created by the currents in the wires. The twisting essentially has the current in the pair of wires going in opposite directions, thus limiting the value of the magnetic field. It does not, however, eliminate the problem of one pair affecting another. Many of the problems associated with twisted wires were solved by the introduction of the more reliable (and much more expensive) coaxial cable (Figure C.50b).

The optical fibre has a series of impressive advantages over the coaxial cable and, of course, twisted wires. These include:

- low attenuation
- no interference from stray electromagnetic signals
- greater capacity (bandwidth), making possible the transmission of very many signals
- security against 'tapping', i.e. unauthorised extraction of information from the signal.

Nature of science

Applied science

The development of optical fibres has been one of the main forces behind the revolution in communications that we experience today. The fast, clear and cheap transfer of information in digital form from one part of the world to another, with all that this implies about the free flow of information and immediate access to it, has a lot to do with the capabilities of modern optical fibres.

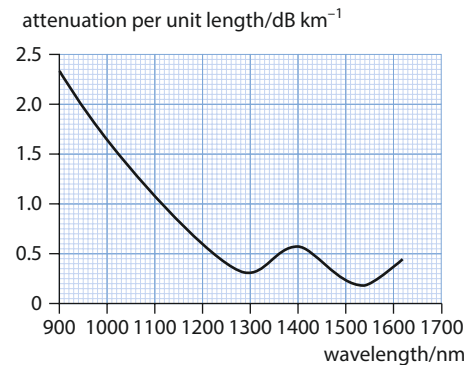


Figure C.49 The variation of specific attenuation with wavelength in a monomode fibre.

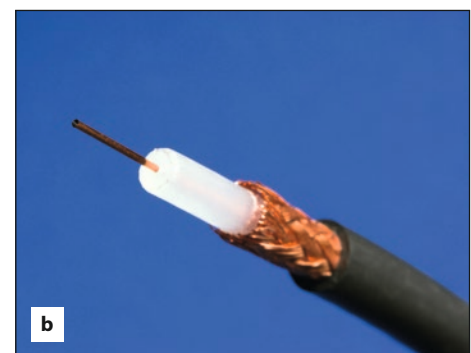
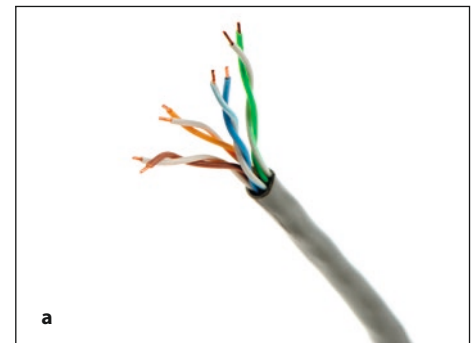
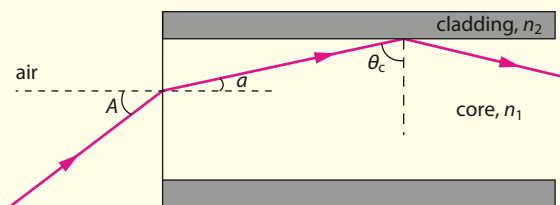


Figure C.50 **a** Twisted wire pairs; **b** coaxial cable.

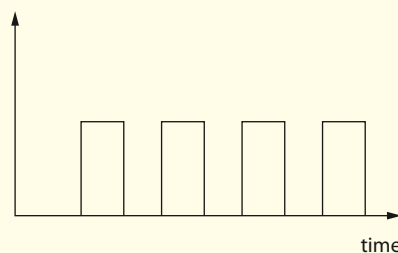
? Test yourself

- 42 Calculate the speed of light in the core of an optical fibre of refractive index 1.45.
- 43 **a** State what is meant by **total internal reflection**.
b Define **critical angle**.
c Explain why total internal reflection can only occur for a ray travelling from a high- to a low-refractive-index medium and not the other way around.
- 44 The refractive indices of the core and the cladding of an optical fibre are 1.50 and 1.46, respectively. Calculate the critical angle at the core-cladding boundary.
- 45 In an optical fibre, n_1 and n_2 are the refractive indices of the core and the cladding, respectively (so $n_1 > n_2$).

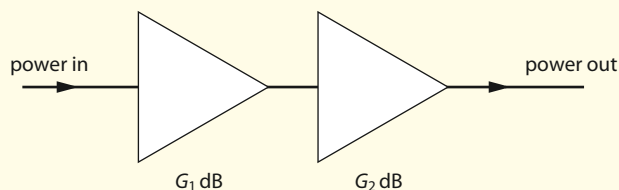


- a** Show that the cosine of the critical angle is given by
- $$\cos \theta_c = \frac{\sqrt{n_1^2 - n_2^2}}{n_1}$$
- b** Hence show that the maximum angle of incidence A from air into the core that will result in the ray being totally internally reflected is given by
- $$A = \arcsin \sqrt{n_1^2 - n_2^2}$$
- c** Calculate the acceptance angle of an optical fibre with a core refractive index of 1.50 and cladding refractive index of 1.40.
- 46 Calculate the acceptance angle of an optical fibre with core and cladding refractive indices equal to 1.52 and 1.44, respectively.
- 47 The refractive index of the cladding of an optical fibre is 1.42. Determine the refractive index of the core such that any ray entering the fibre gets totally internally reflected.
- 48 State **one** crucial property of the glass used in the core of an optical fibre.
- 49 **a** State what is meant by **dispersion** in the context of optical fibres.
b Distinguish between waveguide and material dispersion.

- 50 An optical fibre has a length of 8.00 km. The core of the optical fibre has a refractive index of 1.52 and the core-cladding critical angle is 82° .
a Calculate the speed of light in the core.
b Calculate the minimum and maximum times taken for a ray of light to travel down the length of the fibre.
- 51 The pulse shown below is input into a multimode optical fibre. Suggest the shape of the output pulse after it has travelled a long distance down the fibre.



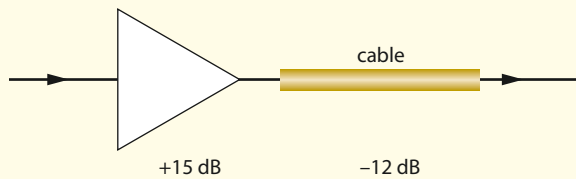
- 52 **a** Distinguish between **monomode** and **multimode** optical fibres.
b Discuss the effect of reducing the fibre core diameter on the bandwidth that can be transmitted by the fibre.
- 53 List **three** advantages of optical fibres in communications.
- 54 State the main cause of attenuation in an optical fibre.
- 55 Two amplifiers of gain G_1 and G_2 (in dB) amplify a signal, as shown below. Calculate the overall gain produced by the two amplifiers.



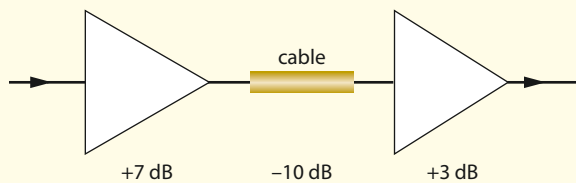
- 56 A signal of power 4.60 mW is attenuated to 3.20 mW. Calculate the power loss in decibels.
- 57 A signal of power 8.40 mW is attenuated to 5.10 mW after travelling 25 km in a cable. Calculate the attenuation per unit length of the cable.
- 58 A coaxial cable has a specific attenuation of 12 dB km^{-1} . The signal must be amplified when the power of the signal falls to 70% of the input power. Determine the distance after which the signal must be amplified.



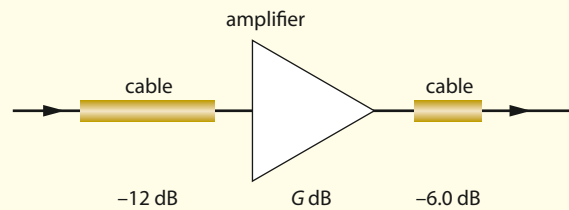
59 A signal is input into an amplifier of gain +15 dB. The signal then travels along a cable, where it suffers a power loss of 12 dB. Calculate the ratio of the output power to the input power.



60 A signal is input into an amplifier of gain +7.0 dB. The signal then travels along a cable, where it suffers a power loss of 10 dB, and is then amplified again by an amplifier of gain +3.0 dB. Calculate the ratio of the output power to the input power.



61 In the arrangement shown below, the output power is twice the input power. Calculate the required gain G of the amplifier.



62 a Sketch a graph (no numbers are required on the axes) to illustrate the variation with wavelength of the specific attenuation in an optical fibre.
b Explain why infrared wavelengths are preferred in optical fibre transmission.

C4 Medical imaging (HL)

This section introduces the use of X-rays and ultrasound in medical imaging. Other imaging techniques, including PET scans and a method based on nuclear magnetic resonance, are also discussed.

C4.1 X-ray imaging

X-rays are electromagnetic radiation with a wavelength around 10^{-10} m. X-rays for medical use are produced in X-ray tubes, in which electrons that have been accelerated to high energies by high potential differences collide with a metal target. As a result of the deceleration suffered by the electrons during the collisions and transitions between energy levels in the target atoms, X-rays are emitted (see Figure C.51). This was the first radiation to be used for medical imaging. Typical hospital X-ray machines operate at voltages of around 15–30 kV for a mammogram or 50–150 kV for a chest X-ray.

X-rays travelling through a medium suffer energy loss, referred to as **attenuation**. The dominant mechanism for this is the photoelectric effect: X-ray photons are absorbed by electrons in the medium and energy is transferred to the electrons. The effect is strongly dependent on the atomic number of the atoms of the medium. There is a substantial difference between the atomic numbers of the elements present in bone ($Z=14$) and soft tissue ($Z=7$), and bone absorbs X-rays more strongly than soft tissue. Hence, an X-ray image will show a **contrast** between bone and soft tissue.

Learning objectives

- Understand the use of X-rays in medical imaging.
- Understand the use of ultrasound in medical imaging.
- Understand magnetic resonance imaging in medicine.

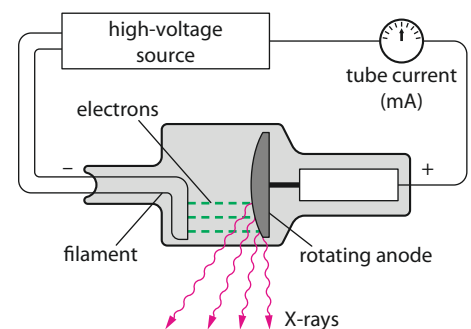


Figure C.51 Schematic diagram of an X-ray tube.

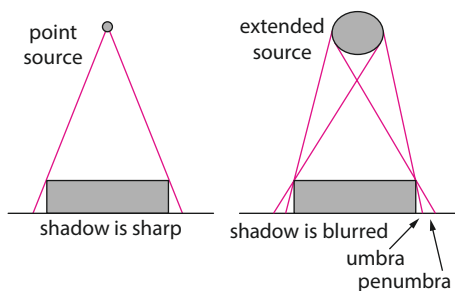


Figure C.52 The shadow cast is blurred if the source is extended.

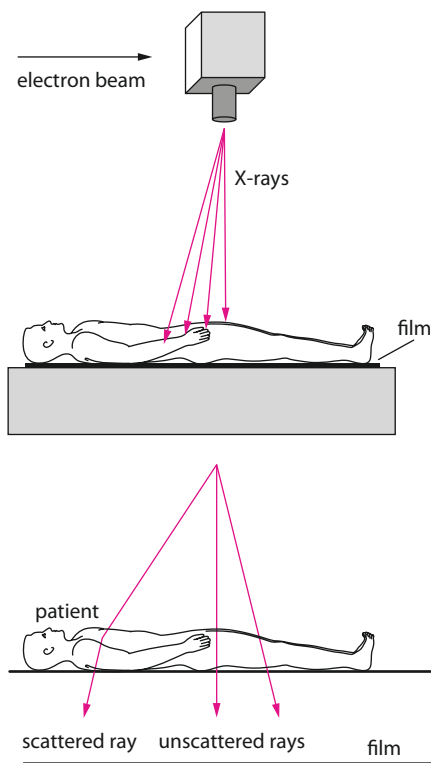


Figure C.53 A beam of X-rays entering a patient. Scattered rays blur the image.

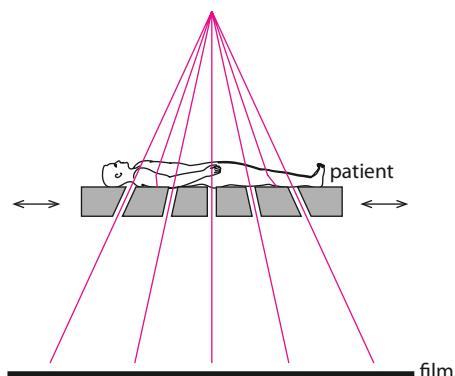


Figure C.54 The use of a grid of lead strips blocks scattered rays, improving the image.

Where there is no substantial difference between the Z numbers of the area to be imaged and of the surrounding area – for example, in the digestive tract – the image can be improved by administering a **contrast medium**. Usually this consists of what is called a **barium meal** (barium sulfate), which the patient swallows. In the intestinal tract, the barium absorbs X-rays more strongly than surrounding tissue.

Those X-rays that pass through a patient's body fall on and expose photographic film. The image created by the X-rays on the film is thus a shadow of the high- Z material against surrounding low- Z tissue.

To increase the **sharpness** of the shadow, the X-ray source should be as point-like as possible (see Figure C.52). The quality of the image is thus improved if the film is as close to the patient as possible, or if the distance from the source to the patient is large. (In the latter case, the intensity of X-rays reaching the patient is diminished, which then requires a longer exposure time.)

The image is also improved if as many scattered rays as possible are prevented from reaching the film (Figure C.53). This can be achieved with the use of a grid of lead strips (lead readily absorbs X-rays) between the patient and the film, as shown in Figure C.54. The strips, about 0.5 mm apart, are closely oriented along the direction of the incoming X-rays, so scattered rays are absorbed. Unwanted images of the lead strips themselves can be minimised by moving the grid sideways back and forth during exposure so that the strip images are blurred.

Lower-energy X-rays tend to be absorbed by the patient's skin and are therefore of little use for imaging. These are usually filtered from the incoming beam.

Because photographic film is much more sensitive to visible light than to X-rays, the exposure time for an X-ray image must be longer. However, this can be significantly reduced by using an intensifying screen, containing fluorescent crystals on the front and rear surfaces and a double-sided photographic film in between. X-rays that have gone through the patient enter this screen and transfer some of their energy to the crystals. This energy is re-emitted as visible light and exposes the film (Figure C.55).

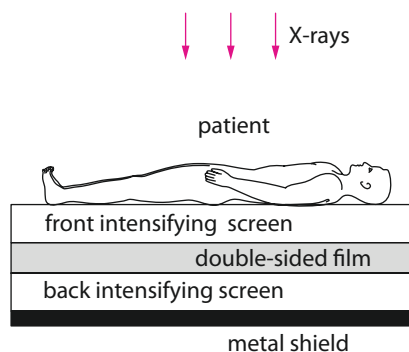


Figure C.55 The use of an intensifying screen increases the brightness of an X-ray image.

A technique called **fluoroscopy** allows for the creation of real-time, dynamic images. X-rays that have passed through the patient fall on a fluorescent screen and visible light is emitted. Directed at a photosurface, these photons cause the emission of electrons, which are accelerated through a potential difference and fall on a second fluorescent screen,



the light from which is fed into a TV monitor. The advantages of a real-time image may, however, be outweighed by the high doses of radiation that are needed.

C4.2 Computed tomography

A major advance in the medical use of X-rays was the development (in 1973) by G. N. Hounsfield and A. Cormack of a technique known as **computed (axial) tomography** (CT) or **computer-assisted tomography** (CAT). This has made possible much more accurate diagnosis using far less invasive procedures, though it does require the use of X-rays. A complete CAT brain scan lasts about 2 s and a whole-body scan about 6 s.

In a whole-body scan, a movable X-ray source emits a beam at right angles to the long axis of the patient, to be detected on the other side. The use of an array of detectors rather than just one reduces exposure levels and the time required for a scan. Figure C.56 shows a view from above the patient's head (the patient is represented by the grey circle).

The source and detectors are rotated around the patient's body and moved along the length of the body. The data can then be combined into a three-dimensional computer image, viewable as two-dimensional 'slices' in any chosen direction and orientation.

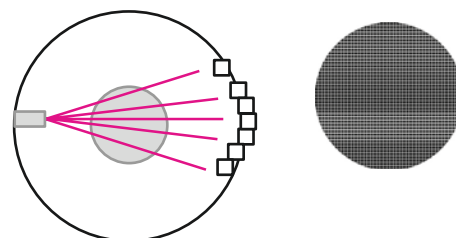


Figure C.56 In a CT scan, an array of detectors records the X-rays passing through the patient in many directions, allowing the construction of a three-dimensional image, which can then be viewed as 'slices' in any chosen direction and orientation.

C4.3 Attenuation

Imagine X-rays of intensity I_0 incident on a medium normally, as shown in Figure C.57. After travelling a distance x through the medium, the intensity of the X-rays has decreased to a value I , given by

$$I = I_0 e^{-\mu x}$$

Here μ is a constant called the **attenuation coefficient**. This coefficient can be determined from the slope of a plot of the logarithm of the intensity versus distance. It depends on the density ρ of the material through which the radiation passes, the atomic number Z of the material and the energy of the X-ray photons.

This relationship is similar to that for radioactive decay, and by analogy we define the **half-value thickness** (HVT), $x_{\frac{1}{2}}$, the penetration distance at which the intensity has been reduced by a factor of 2:

$$\frac{I_0}{2} = I_0 e^{-\mu x_{\frac{1}{2}}}$$

$$\frac{1}{2} = e^{-\mu x_{\frac{1}{2}}}$$

$$\ln \frac{1}{2} = -\mu x_{\frac{1}{2}}$$

Thus the half-value thickness and the attenuation coefficient are related by $\mu x_{\frac{1}{2}} = \ln 2$.

Figure C.58 shows the dependence on energy of the half-value thickness for X-rays and gamma rays in water.

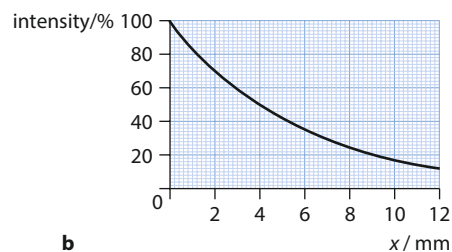
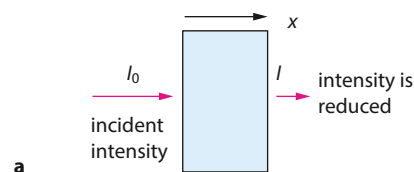


Figure C.57 a Attenuation of radiation in an absorbing medium. b The graph shows the intensity as a function of the penetration depth x .

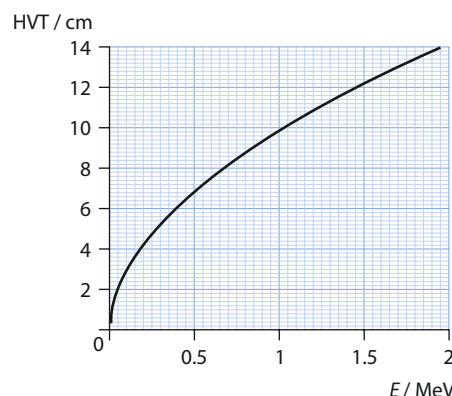


Figure C.58 Half-value thickness as a function of photon energy for X-rays.

Worked example

C.17 A metal sheet of thickness 4.0 mm and half-value thickness 3.0 mm is placed in the path of radiation from a source of X-rays. Calculate the fraction of the source's incident intensity that gets transmitted through the sheet.

We use $I = I_0 e^{-\mu x}$. First we have to find the attenuation coefficient μ . From $\mu x_{\frac{1}{2}} = \ln 2$ we find that

$$\mu = \frac{\ln 2}{x_{\frac{1}{2}}} = \frac{\ln 2}{3.0} = 0.231 \text{ mm}^{-1}.$$

Therefore

$$\begin{aligned} I &= I_0 e^{-\mu x} \\ &= I_0 \exp(-0.231 \times 4.0) \\ &= 0.397 I_0 \end{aligned}$$

or about 40% of the intensity goes through.

Because, for a given atomic number Z and energy E , the attenuation coefficient is proportional to the density, we can define a new coefficient, the **mass absorption coefficient**:

$$\mu_m = \frac{\mu}{\rho}$$

This allows comparisons, for given Z and E , between the attenuation in materials with different densities.

C4.4 Ultrasound

Ultrasound is a major tool in diagnostic medicine. The term refers to sound that is inaudible to the human ear, with a frequency higher than about 20 kHz. The ultrasound used in diagnostic medicine is in the range of about 1–10 MHz. Ultrasound has the advantage over X-rays that it does not deposit radiation damage in the body, and no adverse side effects of its use are known. A disadvantage of ultrasound is that the images are not as detailed as those from X-rays.

The ultrasound is emitted towards the patient's body in short pulses, typically lasting 1 μs , and their reflections off the surfaces of various organs are detected. The idea is thus similar to sonar. A 1 μs pulse of 1 MHz sound contains a single wavelength, while a 1 μs pulse of 10 MHz sound contains 10 wavelengths. The speed of sound in soft tissue is about 1540 m s^{-1} , similar to that in water, giving a wavelength of about 1.54 mm for 1 MHz ultrasound and 0.154 mm for 10 MHz ultrasound.

In general, diffraction places a limit on the size d that can be resolved using a wavelength λ . The constraint is that

$$\lambda < d$$

If resolution of a few millimetres is required, the wavelength used must therefore be less than this. Since 10 MHz ultrasound has a wavelength of about 0.15 mm, in principle it can 'see' objects of this linear size or



resolve objects with this separation. In practice, however, the pulse must contain at least a few full wavelengths for resolution at this level.

For the ultrasound frequencies used in medicine, it is the pulse duration, and not diffraction, that sets the limit on resolution.

The frequency to be used is usually determined by the organ to be studied and the resolution desired. A rough rule of thumb is to use a frequency of $f = 200 \frac{c}{d}$, where c is the speed of sound in tissue and d is the depth of the organ below the body surface. (Thus, for a given frequency, the organ should be at a depth of no more than about 200 wavelengths.)

Worked example

C.18 The stomach is about 10 cm from the body's surface. Suggest what frequency should be used for an ultrasound scan of the stomach.

Applying the formula gives

$$\begin{aligned} f &= 200 \frac{c}{d} \\ &= 200 \times \frac{1548}{0.10} \text{ Hz} \\ &\approx 3 \text{ MHz} \end{aligned}$$

The source of the ultrasound is a **transducer** that converts electrical energy into sound energy, using the phenomenon of **piezoelectricity**. An alternating voltage applied to opposite faces of a crystal such as strontium titanate or quartz will force the crystal to vibrate, emitting ultrasound (see Figure C.59). Likewise, ultrasound falling on such a crystal will produce an alternating voltage at the faces of the crystal. This means that the source of ultrasound can also act as the receiver.

The sound energy must then be directed into the patient's body. In general, when a wave encounters an interface between two different media, part of the wave will be reflected and part will be transmitted. The degree of transmission depends on the **acoustic impedances** of the two media. Acoustic impedance is defined as

$$Z = \rho c$$

where ρ is the density of the medium and c is the speed of sound in that medium. The units of impedance are $\text{kg m}^{-2} \text{ s}^{-1}$. If I_0 is the incident intensity, I_t the transmitted intensity and I_r the reflected intensity,

$$\begin{aligned} \frac{I_t}{I_0} &= \frac{4Z_1 Z_2}{(Z_1 + Z_2)^2} \\ \frac{I_r}{I_0} &= \frac{(Z_1 - Z_2)^2}{(Z_1 + Z_2)^2} \end{aligned}$$

This shows that for most of the energy to be transmitted, the impedances of the two media must be as close to each other as possible (**impedance matching**). The impedance of soft tissue differs from that of air by a factor of about 10^4 , so most ultrasound would be reflected at an air–tissue

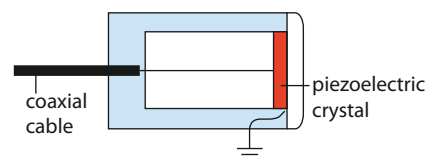


Figure C.59 A piezoelectric crystal vibrates when an ac source is applied to it.

interface. This is why the space between the body and the transducer is filled with a gel-like substance whose impedance closely matches that of soft tissue.

Worked example

C.19 The density of air is 1.2 kg m^{-3} and the speed of sound is 340 m s^{-1} . The density of soft tissue is about 1200 kg m^{-3} and the speed of sound in soft tissue is about 1500 m s^{-1} .

- Calculate the impedances of air and of soft tissue.
- Calculate the fraction of the intensity of ultrasound that would be transmitted from air into soft tissue.
- Comment on your answer.

a The impedances are given by $Z = \rho c$, so $Z_{\text{air}} = 1.2 \times 340 = 408 \text{ kg m}^{-2} \text{ s}^{-1}$ and $Z_{\text{tissue}} = 1200 \times 1500 = 1.8 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}$.

b The transmitted fraction is $\frac{I_t}{I_0} = \frac{4Z_{\text{air}}Z_{\text{tissue}}}{(Z_{\text{air}} + Z_{\text{tissue}})^2} = \frac{4 \times 408 \times 1.8 \times 10^6}{(408 + 1.8 \times 10^6)^2} \approx 9 \times 10^{-4}$

c This is a negligible amount: most of the ultrasound is reflected. This shows the need to place a suitable gel between the transducer and the skin.

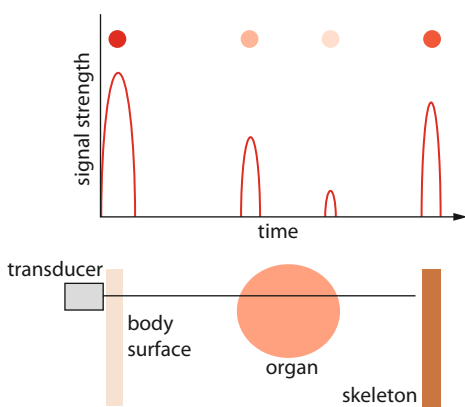


Figure C.60 A beam of ultrasound is directed into the patient. The beam is partially reflected from various organs in the body and the reflected signal is recorded. The dots at the top of the figure represent the strength of the reflected signal.

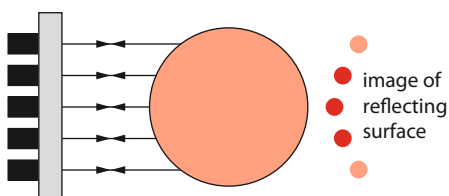


Figure C.61 As the source of ultrasound is moved across the body, a series of dots – whose brightness represents the strength of the reflected signal – is formed. These dots build up an ultrasound image.

In a type of ultrasound scan called an **A scan**, a pulse is directed into the body and the reflected pulses from various interfaces in the body are recorded by the transducer. These signals are converted into electrical energy and the reflected signal is displayed on a cathode-ray oscilloscope (CRO). The CRO signal is, in fact, a graph of signal strength versus time of travel from the transducer to the reflecting surface and back. An example of such a trace is shown in Figure C.60.

The dots in the graph show another way of representing the results, with dot brightness proportional to the signal strength. The A scan provides a one-dimensional image.

Now imagine a series of A scans using parallel beams of ultrasound with a transducer that moves along the body surface or with a series of transducers, as shown in Figure C.61. If these A scans are assembled, the result is a pattern such as the dots on the right of the diagram, which form a two-dimensional image of the surface of the organ.

Now imagine using a series of transducers, each sending one short pulse after the other, with a typical time delay of 1 ms. If these signals are displayed on the CRO screen, the result is a time-varying two-dimensional image of the organ – effectively a real-time ultrasound video. This is called a **B scan**.

Ultrasound can also be put to other uses, including observation of fetal heart movements and measurement of blood-flow velocities. Ultrasound directed at a moving organ or other object gives a reflected signal which is Doppler-shifted (its frequency is shifted). Comparison of the emitted and received frequencies reveals the speed of motion of the reflecting surface.

C4.5 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is based on a phenomenon known as **nuclear magnetic resonance**, and is in many ways superior to CT scans. Unlike CT, the image is constructed without the use of dangerous radiation (despite the term ‘nuclear’), though the procedure is significantly more expensive.

Electrons, protons and some other particles have a property called **spin**. A particle with electric charge and spin behaves like a microscopic magnet (the technical term is **magnetic moment**). In the presence of an external magnetic field, the magnetic moment will align itself either parallel ('spin-up') or anti-parallel ('spin-down') to the direction of the magnetic field. Protons, for example, will have only certain energy values, depending on how their magnetic moments are aligned in the field (see Figure C.62), and the energy difference is proportional to the size of the field.

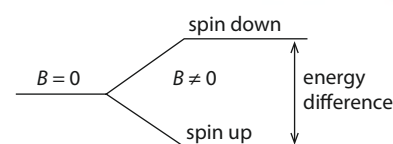


Figure C.62 A normal proton energy level splits in two in an external magnetic field. The spin-up state has a lower energy than the spin-down state.

Exam tip

For exam purposes, it may be helpful to remember the main points of MRI in bullet form:

- A proton aligns itself parallel or anti-parallel to a strong external field.
- A radiofrequency signal forces the proton to change from a spin-up to a spin-down state.
- The proton then returns to the spin-up state, in the process emitting a photon of the same frequency.
- This frequency depends on the energy difference between the states, and therefore on the magnetic field at the proton's position.
- In a region with a different external field, a different radiofrequency will be needed to excite proton transitions.
- A secondary, non-uniform magnetic field is applied so that different parts of the body are exposed to different net magnetic fields.
- Each part of the body is then revealed by a different frequency of emitted photons.
- The rate of these transitions also gives information about tissue type.

The spin-up state has lower energy. If radiofrequency (RF) electromagnetic radiation provides energy to a sample of hydrogen nuclei (i.e. protons) in a magnetic field, those in the spin-up state may absorb photons and make a transition to the higher-energy spin-down state. This will happen if the radiation frequency matches the energy difference between the spin-up and spin-down states (an example of **resonance**). The transition up is followed by a transition down again, with the emission of another photon of the same frequency. Detectors record these photons, and techniques similar to CT scanning are used to create an image; detected photons can be correlated with specific points of emission.

The patient lies in an enclosure surrounded by a powerful magnet which creates a uniform magnetic field (Figure C.63). An additional,

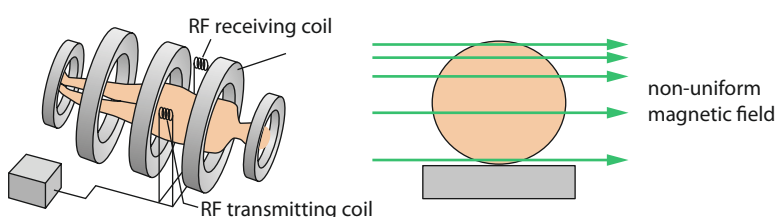


Figure C.63 In MRI, the patient is surrounded by powerful magnets that produce very uniform fields, and by RF coils. The magnetic field is then given a gradient (it is made stronger in some places). Here the field increases as we move upwards.

non-uniform magnetic **gradient field** is superimposed on this, so that the total field varies across the patient (shown in cross-section in the figure). Imagine the variation of the magnetic field to be the same in parallel horizontal planes through the patient.

Now, a magnetic field which is different in different parts of the body will also mean different frequencies of resonant photon absorption and re-emission. Thus, for a particular RF frequency, only one plane within the body will have the correct value of magnetic field for this to take place. Other RF frequencies will give rise to resonance in other parts of the body with different magnetic fields. Variation of the RF frequency will therefore produce a series of photon patterns that can be combined, as with CAT, into a three-dimensional computer image.

More sophisticated techniques also measure the **proton spin relaxation time** – the rate at which excited protons return to their lower states – and these produce images of especially high resolution. Different types of tissue show different relaxation times, thus allowing the identification of particular types of tissue.

The various imaging techniques described in this section are summarised in Table C.1.

Method	Resolution	Advantages	Disadvantages
X-ray	0.5 mm	Inexpensive	Radiation danger Some organs are not accessible Some images are obscured
CT scan	0.5 mm	Can distinguish between different types of tissue	Radiation danger
MRI	1 mm	No radiation dangers Non-invasive Superior images Can distinguish between different types of tissue	Expensive and bulky equipment Difficult for patients who are claustrophobic Long exposures Exposure to magnetic fields difficult for patients with pacemakers and metallic hip implants
Ultrasound	2 mm	No radiation dangers Versatile	Some organs are not accessible

Table C.1 Advantages and disadvantages of various imaging techniques.

Nature of science

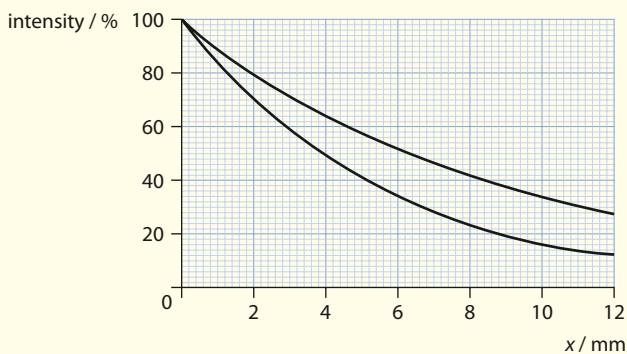
Risk analysis

In many diagnostic imaging techniques, a radiation danger is presented to the patient. That risk must be balanced by the complications to a patient's health should an existing serious condition go unnoticed by not doing the imaging tests. The risk of developing cancer as a result of the radiation dose received in a particular imaging technique depends on the age and gender of the patient, their BMI (body mass index – as a larger dose is needed for imaging thicker body tissues), and the part of the body being investigated. The risks can only be calculated as a probability, based on available evidence. It is believed that there is no completely safe lower limit, and risk increases in proportion to the dose. The doctor's role is to choose the imaging technique that minimises the risk to the patient and provides the greatest overall benefit.

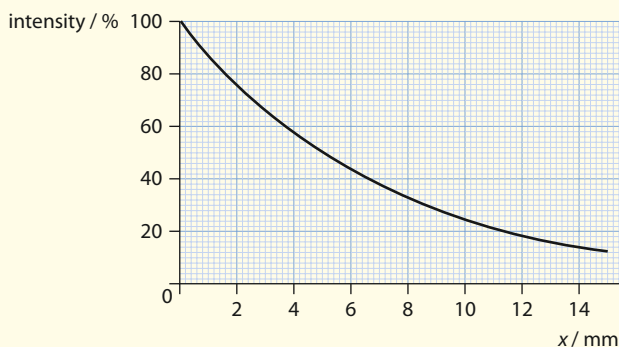


? Test yourself

- 63 a State what is meant by **attenuation**.
 b Describe how X-rays in a medium get attenuated.
- 64 Describe how the exposure time to X-rays during imaging may be reduced.
- 65 Describe the use of contrast media in X-ray imaging.
- 66 X-rays images may be blurry even though the patient is not moving. Suggest how this may be reduced.
- 67 The graph shows the fractions of X-rays of two specific energies transmitted through a thickness x of a sheet of metal.



- a For each energy, determine the half-value thickness for these X-rays in the metal.
- b Determine which graph corresponds to X-rays of higher energy.
- 68 The graph shows the fraction of X-rays of a specific energy transmitted through a thickness x of a sheet of metal.

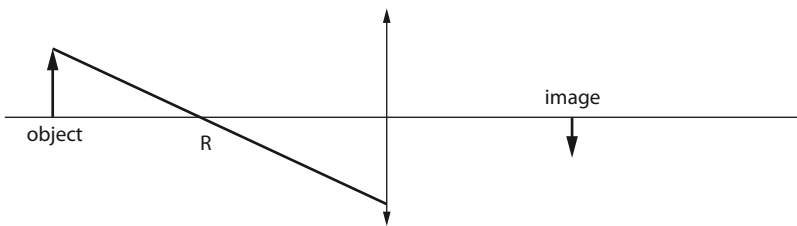


- a Determine the value of the attenuation coefficient for these X-rays in this metal.
- b Determine the thickness of metal that is required to reduce the transmitted intensity by 80%.

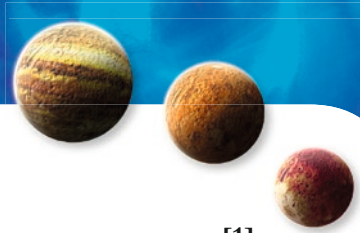
- 69 A piece of metal 4.0 mm thick reduces the intensity of X-rays passing through it by 40%. What thickness of the same metal is required to reduce the intensity by 80%?
- 70 The half-value thickness for a beam of X-rays in a particular metal is 3.0 mm. Determine the fraction of X-ray intensity that is transmitted through 1.0 mm of this metal.
- 71 The X-rays used in medicine are usually not monoenergetic (i.e. of a single energy). It is said that these beams become 'harder' as they are allowed to pass through a material.
- a Suggest what is meant by this statement and why it is true.
- b The half-value thickness of a certain absorber for X-rays of energy 20 keV is 2.2 mm and that for 25 keV X-rays is 2.8 mm. A beam containing equal quantities of X-rays of these two energies is incident on 5.0 mm of the absorber. Calculate the ratio of 25 keV to 20 keV photons that are transmitted.
- 72 State what is meant by **ultrasound** and describe how it is produced.
- 73 Estimate the resolution that can be achieved with ultrasound of frequency 5 MHz. (Take the speed of sound in soft tissue to be 1540 m s^{-1} .)
- 74 a State what is meant by **impedance**.
 b The density of muscle is 940 kg m^{-3} and the impedance of muscle is $1.4 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}$. Calculate the speed of sound in muscle.
- 75 Ultrasound is directed from air into a type of tissue. The impedance of air is $420 \text{ kg m}^{-2} \text{ s}^{-1}$ and that of tissue is $1.6 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}$.
- a Calculate the fraction of the incident intensity that gets transmitted into the tissue.
- b Comment on your answer.
- 76 Distinguish between **A scans** and **B scans** in ultrasound imaging.
- 77 Suggest the role of the gradient magnetic field in MRI.
- 78 Describe the method of **magnetic resonance imaging**.

Exam-style questions

- 1 a State what is meant by the **focal length** of a converging lens. [2]
- b In order to view the detail on an ancient coin, an art dealer holds a converging lens 2.0 cm above the coin. A virtual upright image of the coin is formed with a magnification of 5.0. Calculate the focal length of the lens [3]
- c Determine where the object should be placed so that the magnification produced is as large as possible. [3]
- 2 The diagram below shows an object placed in front of a converging lens. The lens forms an image of the object. The diagram also shows a ray R from the object.



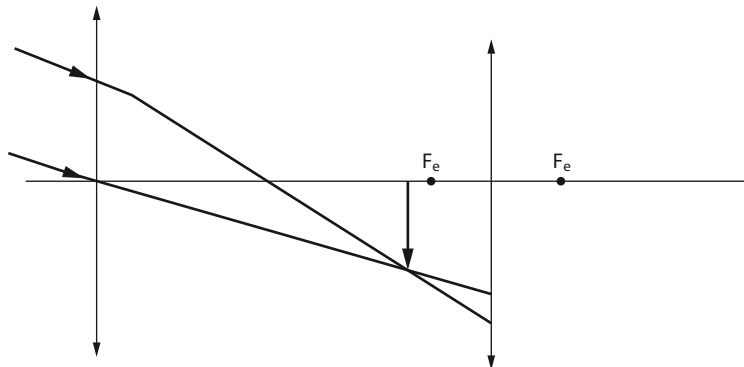
- a On a copy of the diagram, extend the ray R to show how it refracts in the lens. [1]
- b Draw an appropriate ray to locate the focal points of the lens. [2]
- c State and explain whether the image formed is real or virtual. [2]
- d The upper half of the lens is covered with opaque paper. State and explain the effect of this, if any, on the image. [2]
- e A converging lens of focal length 4.0 cm is used as a magnifying glass in order to view an object of length 5.0 mm placed at right angles to the axis of the lens. The image is formed 25 cm from the eye, which is placed very close to the lens. Determine:
- i the distance of the object from the lens [2]
 - ii the length of the image [2]
 - iii the angle subtended by the image at the eye. [1]
- 3 A compound microscope consists of an objective lens of focal length 15 mm and an eyepiece lens of focal length 60 mm. The final image of an object placed 20 mm from the objective is formed 25 cm from the eyepiece lens.
- a Determine:
- i the distance of the image formed by the objective from the objective lens [2]
 - ii the distance of the image in i from the eyepiece lens. [2]
- b i State what is meant by the **angular magnification** of a microscope. [2]
- ii Determine the angular magnification of the microscope. [2]
- c The object has a length of 8.0 mm and is placed at right angles to the axis of the microscope. Calculate:
- i the length of the final image [1]
 - ii the angle subtended by the final image at the eyepiece lens. [1]



4 An astronomical refracting telescope consists of two converging lenses.

a Suggest a reason why the diameter of the objective lens of a telescope should be large. [1]

b An astronomical telescope is used to view the Sun. The diagram below (not to scale) shows the formation of the intermediate image of the Sun. On a copy of the diagram, draw lines to show the formation of the real image of the Sun in the eyepiece. [2]



c The focal length of the objective is 1.00 m and that of the eyepiece is 0.10 m. Calculate the distance of the image in **b** from the eyepiece, given that the eyepiece forms a **real** image 0.455 m from the eyepiece. [2]

d The rays of the Sun make an angle of 0.055 rad with the axis of the objective. Determine the size of the image in the eyepiece. [3]

5 a An object of height 3.0 cm is placed 8.0 cm in front of a concave spherical mirror of focal length 24 cm. Calculate:

i the position of the image [1]

ii the height of the image. [2]

b Draw a ray diagram to illustrate your answers to **i** and **ii**. [3]

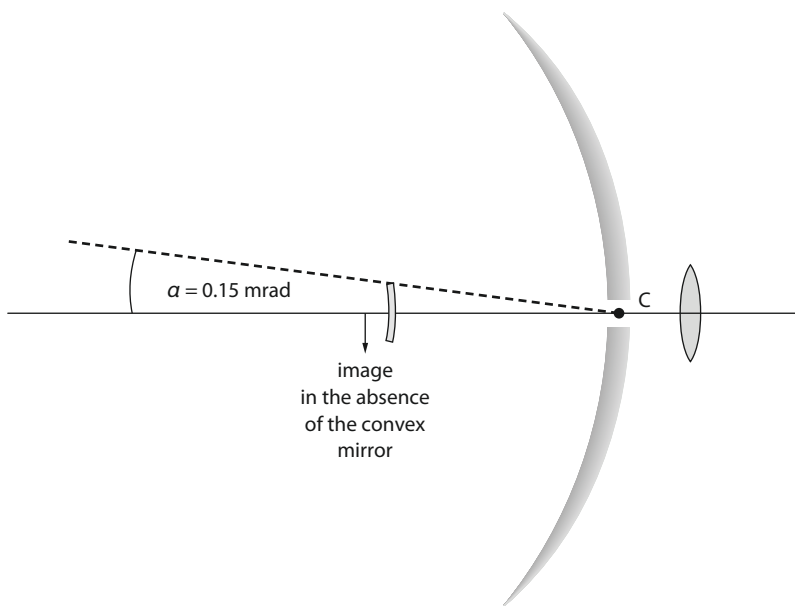
c An object is placed in front of a mirror. An upright image half the height of the object is formed behind the mirror. The distance between the object and the image is 120 cm. Calculate the focal length of the mirror. [3]

d Telescopes use mirrors rather than lenses.

i Outline **two** advantages of mirrors over lenses in a telescope. [4]

ii State **one** advantage of parabolic mirrors over spherical mirrors in a telescope. [1]

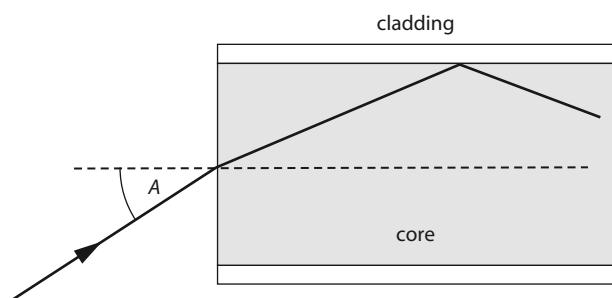
- 6 The diagram below (not to scale) shows a Cassegrain-type reflecting telescope. The small arrow shows the image of a planetary feature that would be formed by the concave mirror in the absence of the small convex mirror. The actual feature subtends an angle $\alpha = 1.50 \times 10^{-4}$ rad at the concave mirror. The focal length of the concave mirror is 10.0 m.



- a Calculate the length of the image shown here. [2]
- b This image serves in turn as the object for the small convex mirror, which produces a real image at C. The concave and convex mirrors are separated by 9.00 m. Calculate:
- i the focal length of the convex mirror [2]
 - ii the magnification of the convex mirror [1]
 - iii the height of the image at C. [1]
- c The image at C is viewed through a converging lens of focal length 12 cm, forming a virtual image very far away.
- i Calculate the angle subtended at the converging lens by the image at C. [2]
 - ii Hence calculate the overall angular magnification of this system. [2]
- 7 a State what is meant by a **lens aberration**. [1]
- b i **Spherical** and **chromatic aberration** are two common types of lens aberration. Describe what is meant by each. [4]
- ii Describe **one** way in which each of the aberrations in i may be reduced. [2]
- 8 Two objects, A and B, each of height 5.0 cm, are placed in front of a concave mirror of focal length 24 cm. The distances of the objects A and B from the mirror are, respectively, 40 cm and 30 cm.
- a Calculate:
- i the distance between the images of A and B [3]
 - ii the difference in heights of the images of A and B. [3]
- b A rod of length 10 cm is placed in front of the concave mirror such that it is parallel to the principal axis of the mirror and 5.0 cm to the side of it. The front of the rod is 30 cm from the mirror. Use your answer in a to determine whether the image of the rod:
- i has the same length as the rod itself [2]
 - ii is parallel to the principal axis. [2]



- 9 The diagram shows a ray of light entering the core of an optical fibre from air. The core has a refractive index of 1.58 and the cladding a refractive index of 1.45.



- a Determine:
- the critical angle at the core–cladding boundary [2]
 - the largest angle of incidence A such that the ray will undergo total internal reflection at the core–cladding boundary. [2]
- b Distinguish between **waveguide dispersion** and **material dispersion** in an optical fibre. [3]
- c Outline how **i** waveguide dispersion and **ii** material dispersion may be reduced. [4]
- d The power of a signal input into an optical fibre is 25.0 mW. The attenuation per unit length for this fibre is 3.50 dB km⁻¹. The signal power must not fall below 15.0 μW.
- State one source of attenuation in an optical fibre. [1]
 - Determine the distance after which the signal must be amplified. [3]
- HL** 10 a **i** State what is meant by **ultrasound**. [1]
- ii** Ultrasound and X-rays are equally capable of imaging parts of the body. Suggest why ultrasound would be the preferred method of imaging. [1]
- b The impedance of air is $Z_{\text{air}} = 410 \text{ kg m}^{-2} \text{ s}^{-1}$ and that of soft tissue is $Z_{\text{tissue}} = 1.8 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}$.
The fraction of the intensity that gets reflected back from the air–tissue boundary is $\frac{I_r}{I_0} = \frac{(Z_{\text{air}} - Z_{\text{tissue}})^2}{(Z_{\text{air}} + Z_{\text{tissue}})^2}$.
- Calculate this fraction. [1]
 - Comment on the answer to **i**, suggesting a solution to the problem it reveals. [3]
- c A pulse of ultrasound is reflected from the boundary of an organ 6.5 ms after it is emitted. The region between the surface of the skin, where the pulse originates, and the organ is filled with tissue in which the speed of sound is 1500 ms⁻¹. Estimate the distance of the organ boundary from the surface of the skin. [2]
- HL** 11 a State what is meant by **half-value thickness** (HVT). [1]
- b The half-value thickness of soft tissue for X-rays of a given energy is 4.10 mm.
- After a distance x in soft tissue, the fraction of the incident intensity of X-rays that gets transmitted is 0.650. Determine this distance. [3]
 - State and explain the effect, if any, on the answer to **i** if X-rays with a larger half-value thickness were to be used. [2]
- c Outline how, in X-ray imaging, the following are achieved:
- reduction of the blurring in the image [2]
 - reduction of the exposure time. [2]
- HL** 12 Outline the technique of **magnetic resonance imaging**. [6]

Option D Astrophysics

D1 Stellar quantities

This section begins with a brief description of the various objects that comprise the universe, especially stars. We discuss astronomical distances and the main characteristics of stars: their luminosity and apparent brightness. Table **D.1** presents a summary of key terms.

D1.1 Objects in the universe

We live in a part of space called the **solar system**: a collection of eight major planets (Mercury, Venus, the Earth, Mars, Jupiter, Saturn, Uranus and Neptune) bound in elliptical orbits around a star we call the Sun. Pluto has been stripped of its status as a major planet and is now called a ‘dwarf planet’. The orbit of the Earth is almost circular; that of Mercury is the most elliptical. All planets revolve around the Sun in the same direction. This is also true of the comets, with a few exceptions, the most famous being Halley’s comet. All the planets except Mercury and Venus have moons orbiting them.

Leaving the solar system behind, we enter interstellar space, the space between stars. At a distance of 4.2 light years (a light year is the distance travelled by light in one year) we find Proxima Centauri, the nearest star to us after the Sun. Many stars find themselves in **stellar clusters**, groupings of large numbers of stars that attract each other gravitationally and are relatively close to one another. Stellar clusters are divided into two groups: **globular clusters**, containing large numbers of mainly old, evolved stars, and **open clusters**, containing smaller numbers of young stars (some are very hot) that are further apart, Figure **D.1**.

Very large numbers of stars and stellar clusters (about 200 billion of them) make up our **galaxy**, the Milky Way, a huge assembly of stars that are kept together by gravity. A galaxy with spiral arms (similar to the one in Figure **D.2a**), it is about 120 000 light years across; the arm in which our solar system is located can be seen on a clear dark night as the spectacular ‘milky’ glow of millions of stars stretching in a band across the sky.

As we leave our galaxy behind and enter intergalactic space, we find that our galaxy is part of a group of galaxies – a **cluster** (such as the one shown in Figure **D.2b**), known as the Local Group. There are about 30 galaxies in the Local Group, the nearest being the Large Magellanic

Learning objectives

- Describe the main objects comprising the universe.
- Describe the nature of stars.
- Understand astronomical distances.
- Work with the method of parallax.
- Define luminosity and apparent brightness and solve problems with these quantities and distance.

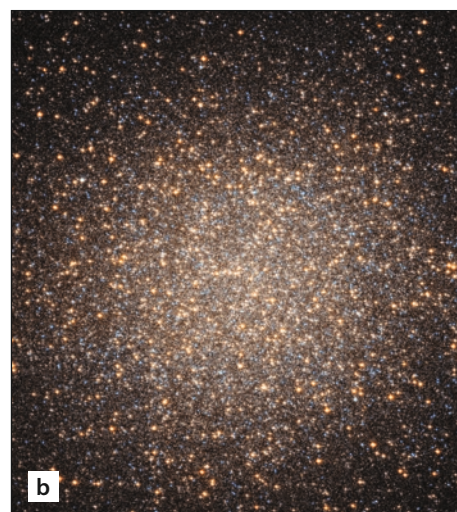


Figure D.1 a The open cluster M36; b the globular cluster M13.

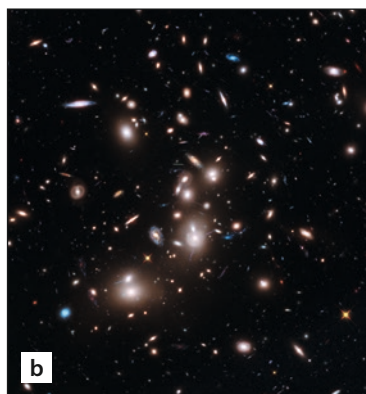


Figure D.2 a The spiral galaxy M74; b the galaxy cluster Abell 2744.

Cloud at a distance of about 160 000 light years. In this group, we also find the Andromeda galaxy, a spiral galaxy like our own and the largest member of the Local Group. Andromeda is expected to collide with the Milky Way in 4 billion years or so.

As we move even further out, we encounter collections of clusters of galaxies, known as **superclusters**. If we look at the universe on a really large scale, more than 10^8 light years, we then see an almost uniform distribution of matter. At such enormous scales, every part of the universe looks the same.

Binary star	Two stars orbiting a common centre
Black dwarf	The remnant of a white dwarf after it has cooled down. It has very low luminosity
Black hole	A singularity in space time; the end result of the evolution of a very massive star
Brown dwarf	Gas and dust that did not reach a high enough temperature to initiate fusion. These objects continue to compact and cool down
Cepheid variable	A star of variable luminosity. The luminosity increases sharply and falls off gently with a well-defined period. The period is related to the absolute luminosity of the star and so can be used to estimate the distance to the star
Cluster of galaxies	Galaxies close to one another and affecting one another gravitationally, behaving as one unit
Comet	A small body (mainly ice and dust) orbiting the Sun in an elliptical orbit
Constellation	A group of stars in a recognisable pattern that <i>appear</i> to be near each other in space
Dark matter	Generic name for matter in galaxies and clusters of galaxies that is too cold to radiate. Its existence is inferred from techniques other than direct visual observation
Galaxy	A collection of a very large number of stars mutually attracting one another through the gravitational force and staying together. The number of stars in a galaxy varies from a few million in dwarf galaxies to hundreds of billions in large galaxies. It is estimated that 100 billion galaxies exist in the observable universe
Interstellar medium	Gases (mainly hydrogen and helium) and dust grains (silicates, carbon and iron) filling the space between stars. The density of the interstellar medium is very low. There is about one atom of gas for every cubic centimetre of space. The density of dust is a trillion times smaller. The temperature of the gas is about 100 K
Main-sequence star	A normal star that is undergoing nuclear fusion of hydrogen into helium. Our Sun is a typical main-sequence star
Nebula	Clouds of 'dust', i.e. compounds of carbon, oxygen, silicon and metals, as well as molecular hydrogen, in the space in between stars
Neutron star	The end result of the explosion of a red supergiant; a very small star (a few tens of kilometres in diameter) and very dense. This is a star consisting almost entirely of neutrons. The neutrons form a superfluid around a core of immense pressure and density. A neutron star is an astonishing macroscopic example of microscopic quantum physics
Planetary nebula	The ejected envelope of a red giant star
Red dwarf	A very small star with low temperature, reddish in colour
Red giant	A main-sequence star evolves into a red giant – a very large, reddish star. There are nuclear reactions involving the fusion of helium into heavier elements
Stellar cluster	A group of stars that are physically near each other in space, created by the collapse of a single gas cloud
Supernova (Type Ia)	The explosion of a white dwarf that has accreted mass from a companion star exceeding its stability limit
Supernova (Type II)	The explosion of a red supergiant star: The amount of energy emitted in a supernova explosion can be staggering – comparable to the total energy radiated by our Sun in its entire lifetime!
White dwarf	The end result of the explosion of a red giant. A small, dense star (about the size of the Earth), in which no nuclear reactions take place. It is very hot but its small size gives it a very low luminosity

Table D.1 Definitions of terms.



Worked example

D.1 Take the density of interstellar space to be one atom of hydrogen per cm^3 of space. How much mass is there in a volume of interstellar space equal to the volume of the Earth? Give an order-of-magnitude estimate without using a calculator.

The volume of the Earth is

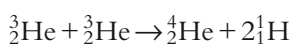
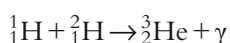
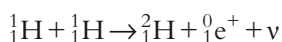
$$\begin{aligned} V &\approx \frac{4}{3}\pi R^3 \\ &\approx \frac{4}{3} \times 3 \times (6 \times 10^6)^3 \text{ m}^3 \\ &\approx 4 \times 200 \times 10^{18} \\ &\approx 10^{21} \text{ m}^3 \end{aligned}$$

The number of atoms in this volume is $10^{21} \times 10^6 = 10^{27}$ atoms of hydrogen (one atom in a cubic cm implies 10^6 atoms in a cubic metre). This corresponds to a mass of

$$10^{27} \times 10^{-27} \text{ kg} \approx 1 \text{ kg}.$$

D1.2 The nature of stars

A star such as our own Sun radiates an enormous amount of power into space – about 10^{26} J s^{-1} . The source of this energy is nuclear fusion in the interior of the star, in which nuclei of hydrogen fuse to produce helium and energy. Because of the **high temperatures** in the interior of the star, the electrostatic repulsion between protons can be overcome, allowing hydrogen nuclei to come close enough to each other to fuse. Because of the **high pressure** in stellar interiors, the nuclei are sufficiently close to each other to give a high probability of collision and hence fusion. The sequence of nuclear fusion reactions that take place is called the **proton–proton cycle** (Figure D.3):



The net result of these reactions is that four hydrogen nuclei turn into one helium nucleus (to see this multiply the first two reactions by 2 and add side by side). Energy is released at each stage of the cycle, but most of it is released in the third and final stage. The energy produced is carried away by the photons (and neutrinos) produced in the reactions. As the photons move outwards they collide with the surrounding material, creating a **radiation pressure** that opposes the gravitational pressure arising from the mass of the star. In the outer layers, convection currents also carry the energy outwards. In this way, the balance between radiation and gravitational pressures keeps the star in equilibrium (Figure D.4).

Nuclear fusion provides the energy that is needed to keep the star hot, so that the radiation pressure is high enough to oppose further gravitational contraction, and at the same time to provide the energy that the star is radiating into space.

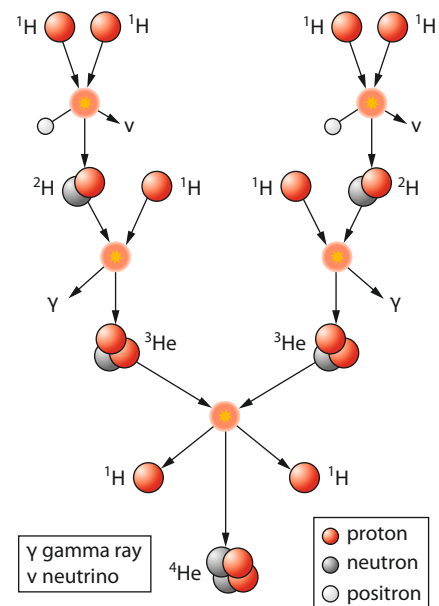


Figure D.3 The proton–proton cycle of fusion reactions.

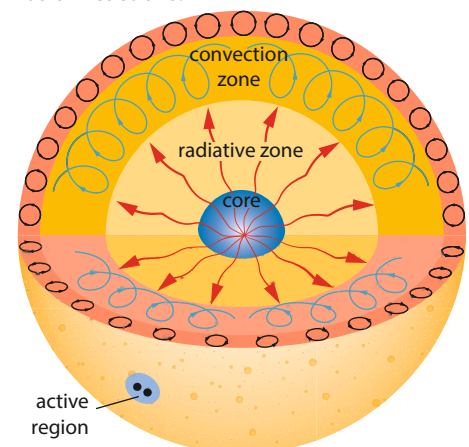


Figure D.4 The stability of a star depends on equilibrium between two opposing forces: gravitation, which tends to collapse the star, and radiation pressure, which tends to make it expand.

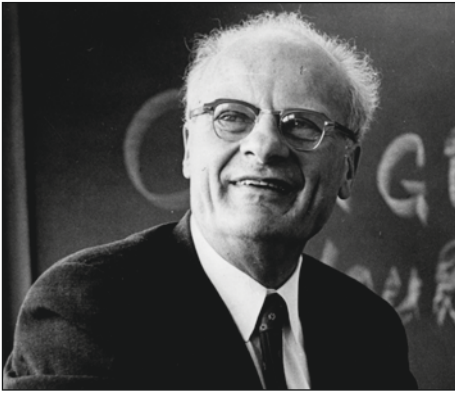


Figure D.5 Hans Bethe, who unravelled the secrets of energy production in stars.

Many of the details of nuclear fusion reactions in stars were worked out by the legendary Cornell physicist Hans Bethe (1906–2005) (Figure D.5).

D1.3 Astronomical distances

In astrophysics, it is useful to have a more convenient unit of distance than the metre!

A **light year** (ly) is the distance travelled by light in one year. Thus:

$$\begin{aligned} 1 \text{ ly} &= 3 \times 10^8 \times 365 \times 24 \times 60 \times 60 \text{ m} \\ &= 9.46 \times 10^{15} \text{ m} \end{aligned}$$

Also convenient for measuring large distances is the **parsec** (pc), a unit that will be properly defined in Section D1.4:

$$1 \text{ pc} = 3.26 \text{ ly} = 3.09 \times 10^{16} \text{ m}$$

Yet another convenient unit is the **astronomical unit** (AU), which is the average radius of the Earth's orbit around the Sun:

$$1 \text{ AU} = 1.5 \times 10^{11} \text{ m}$$

Exam tip

$$1 \text{ AU} = 1.5 \times 10^{11} \text{ m}$$

$$1 \text{ ly} = 9.46 \times 10^{15} \text{ m}$$

$$1 \text{ pc} = 3.09 \times 10^{16} \text{ m} = 3.26 \text{ ly}$$

The average distance between stars in a galaxy is about 1 pc. The distance to the nearest star (Proxima Centauri) is approximately $4.2 \text{ ly} = 1.3 \text{ pc}$. A simple message sent to a civilisation on Proxima Centauri would thus take 4.2 yr to reach it and an answer would arrive on Earth another 4.2 yr later.

The average distance between galaxies varies from about 100 kiloparsecs (kpc) for galaxies within the same cluster to a few megaparsecs (Mpc) for galaxies belonging to different clusters.

Worked examples

D.2 The Local Group is a cluster of some 30 galaxies, including our own Milky Way and the Andromeda galaxy. It extends over a distance of about 1 Mpc. Estimate the average distance between the galaxies of the Local Group.

Assume that a volume of

$$\begin{aligned} V &\approx \frac{4}{3}\pi R^3 \\ &\approx \frac{4}{3} \times 3 \times (0.5)^3 \text{ Mpc}^3 \\ &\approx 0.5 \text{ Mpc}^3 \end{aligned}$$

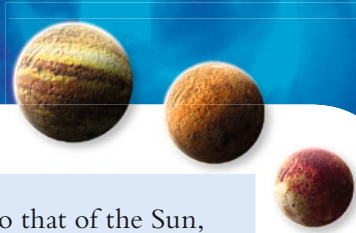
is uniformly shared by the 30 galaxies. Then to each there corresponds a volume of

$$\frac{0.5}{30} \text{ Mpc}^3 = 0.017 \text{ Mpc}^3$$

The linear size of each volume is thus

$$\begin{aligned} \sqrt[3]{0.017 \text{ Mpc}^3} &\approx 0.3 \text{ Mpc} \\ &= 300 \text{ kpc} \end{aligned}$$

so we may take the average separation of the galaxies to be about 300 kpc.



D.3 The Milky Way galaxy has about 2×10^{11} stars. Assuming an average stellar mass equal to that of the Sun, estimate the mass of the Milky Way.

The mass of the Sun is 2×10^{30} kg, so the Milky Way galaxy has a mass of about $2 \times 10^{11} \times 2 \times 10^{30}$ kg = 4×10^{41} kg

D.4 The observable universe contains some 100 billion galaxies. Assuming an average galactic mass comparable to that of the Milky Way, estimate the mass of the observable universe.

The mass is $100 \times 10^9 \times 4 \times 10^{41}$ kg = 4×10^{52} kg.

D1.4 Stellar parallax and its limitations

The **parallax** method is a means of measuring astronomical distances. It takes advantage of the fact that, when an object is viewed from two different positions, it appears displaced relative to a fixed background. If we measure the angular position of a star and then repeat the measurement some time later, the two positions will be different relative to a background of very distant stars, because in the intervening time the Earth has moved in its orbit around the Sun. We make two measurements of the angular position of the star six months apart; see Figure D.6.

The distance between the two positions of the Earth is $D = 2R$, the diameter of the Earth's orbit around the Sun ($R = 1.5 \times 10^{11}$ m). The distance d to the star is given by

$$\tan p = \frac{R}{d} \Rightarrow d = \frac{R}{\tan p}$$

Since the parallax angle is very small, $\tan p \approx p$, where the parallax p is measured in radians, and so $d = \frac{R}{p}$.

The parallax angle (shown in Figure D.6) is the angle, at the position of the star, that is subtended by a distance equal to the radius of the Earth's orbit around the Sun (1 AU).

The parallax method can be used to define a common unit of distance in astronomy, the **parsec**. One parsec (from **parallax second**) is the distance to a star whose parallax is 1 arc second, as shown in Figure D.7. An arc second is $1/3600$ of a degree.

In conventional units,

$$1 \text{ pc} = \frac{1 \text{ AU}}{1''} = \frac{1.5 \times 10^{11}}{\left(\frac{2\pi}{360}\right) \left(\frac{1}{3600}\right)} \text{ m} = 3.09 \times 10^{16} \text{ m}$$

(The factor of $\frac{2\pi}{360}$ converts degrees to radians.)

If the parallax of a star is known to be p arc seconds, its distance is

$$d \text{ (in parsecs)} = \frac{1}{p} \text{ (in arc seconds)}.$$

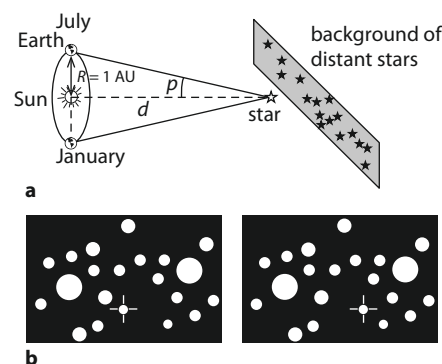


Figure D.6 **a** The parallax of a star. **b** Two 'photographs' of the same region of the sky taken six months apart. The position of the star (indicated by a cross) has shifted, relative to the background stars, in the intervening six months.

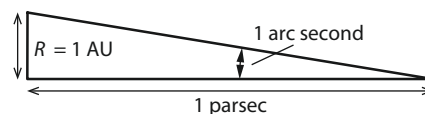


Figure D.7 Definition of a parsec: 1 parsec is the distance at which 1 AU subtends an angle of 1 arc second.

Exam tip

You will not be asked to provide these derivations in an exam. You should just know that d (in parsecs) = $\frac{1}{p}$ (in arc seconds).

You must also understand the limitations of this method.



Ancient methods are still useful!

Astrophysicists still use an ancient method of measuring the apparent brightness of stars. This is the **magnitude scale**. Given a star of apparent brightness b , we assign to that star an **apparent magnitude** m , defined by

$$\frac{b}{b_0} = 100^{-m/5}$$

The value $b_0 = 2.52 \times 10^{-8} \text{ W m}^{-2}$ is taken as the reference value for apparent brightness. Taking logarithms (to base 10) gives the equivalent form

$$m = -\frac{5}{2} \log\left(\frac{b}{b_0}\right)$$

Since $100^{1/5} = 2.512$, the first equation above can also be written as

$$\frac{b}{b_0} = 2.512^{-m}$$

If the star is too far away, however, the parallax angle is too small to be measured and this method fails. Typically, measurements from observatories on Earth allow distances up to 100 pc to be determined by the parallax method, which is therefore mainly used for nearby stars. Using measurements from satellites without the distortions caused by the Earth's atmosphere (turbulence, and variations in temperature and refractive index), much larger distances can be determined using the parallax method. The Hipparcos satellite (launched by ESA, the European Space Agency, in 1989) measured distances to stars 1000 pc away; Gaia, launched by ESA in December 2013, is expected to do even better, extending the parallax method to distances beyond 100 000 pc!

Table D.2 shows the five nearest stars (excluding the Sun).

Star	Distance/ly
Proxima Centauri	4.3
Barnard's Star	5.9
Wolf 359	7.7
Lalande 21185	8.2
Sirius	8.6

Table D.2 Distances to the five nearest stars.

D1.5 Luminosity and apparent brightness

Stars are assumed to radiate like black bodies. For a star of surface area A and absolute surface temperature T , we saw in Topic 8 that the power radiated is

$$L = \sigma AT^4$$

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

The power radiated by a star is known in astrophysics as the **luminosity**. It is measured in watts.



Consider a star of luminosity L . Imagine a sphere of radius d centred at the location of the star. The star radiates uniformly in all directions, so the energy radiated in 1 s can be thought of as distributed over the surface of this imaginary sphere. A detector of area a placed somewhere on this sphere will receive a small fraction of this total energy (see Figure D.8a).

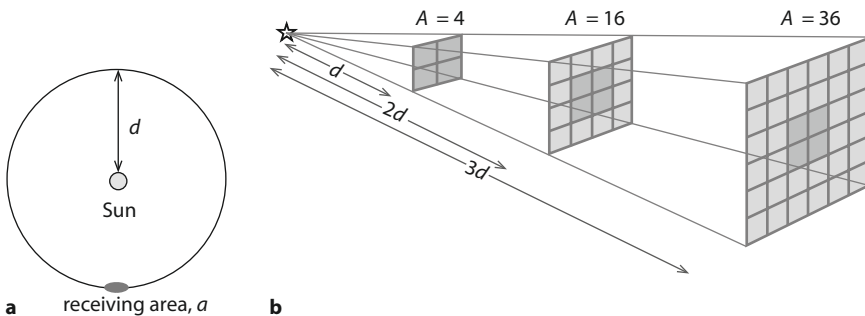


Figure D.8 **a** The Sun's energy is distributed over an imaginary sphere of radius equal to the distance between the Sun and the observer. The observer thus receives only a very small fraction of the total energy, equal to the ratio of the receiver's area to the total area of the imaginary sphere. **b** The inverse square law.

This fraction is equal to the ratio of the detector area a to the total surface area of the sphere; that is, the received power is $\frac{aL}{4\pi d^2}$.

This shows that the apparent brightness is directly proportional to the luminosity, and varies as the inverse square of the star's distance (see Figure D.8b). By combining the formula for luminosity with that for apparent brightness, we see that

$$b = \frac{\sigma AT^4}{4\pi d^2}$$

Apparent brightness is easily measured (with a charge-coupled device, or CCD). If we also know the distance to a star, then we can determine its luminosity. Knowing the luminosity of a star is important because it tells a lot about the nature of the star.

Exam tip

In many problems you will need to know that the surface area of a sphere of radius R is $A = 4\pi R^2$.

The received power per unit area is called the **apparent brightness** and is given by

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

The unit of apparent brightness is W m^{-2} .

Exam tip

Apparent brightness in astrophysics is what is normally called intensity in physics.

Worked examples

D.5 The radius of star A is three times that of star B and its temperature is double that of B. Find the ratio of the luminosity of A to that of B.

$$\begin{aligned} \frac{L_A}{L_B} &= \frac{\sigma 4\pi (R_A)^2 T_A^4}{\sigma 4\pi (R_B)^2 T_B^4} \\ &= \frac{(R_A)^2 T_A^4}{(R_B)^2 T_B^4} \\ &= \frac{(3R_B)^2 (2T_B)^4}{(R_B)^2 T_B^4} \\ &= 3^2 \times 2^4 = 144 \end{aligned}$$

D.6 The stars in Worked example **D.5** have the same apparent brightness when viewed from the Earth. Calculate the ratio of their distances.

$$\begin{aligned}\frac{b_A}{b_B} &= 1 \\ &= \frac{L_A/(4\pi d_A^2)}{L_B/(4\pi d_B^2)} \\ &= \frac{L_A d_B^2}{L_B d_A^2} \\ \Rightarrow \frac{d_A}{d_B} &= 12\end{aligned}$$

D.7 The apparent brightness of a star is $6.4 \times 10^{-8} \text{ W m}^{-2}$. Its distance is 15 ly. Find its luminosity.

We use $b = \frac{L}{4\pi d^2}$ to find

$$\begin{aligned}L &= b4\pi d^2 = \left(6.4 \times 10^{-8} \frac{\text{W}}{\text{m}^2}\right) \times 4\pi \times (15 \times 9.46 \times 10^{15})^2 \text{ m}^2 \\ &= 1.6 \times 10^{28} \text{ W}\end{aligned}$$

D.8 A star has half the Sun's surface temperature and 400 times its luminosity. Estimate the ratio of the star's radius to that of the Sun. The symbol R_\odot stands for the radius of the Sun.

We have that

$$\begin{aligned}400 &= \frac{L}{L_\odot} = \frac{\sigma 4\pi(R^2)T^4}{\sigma 4\pi(R_\odot)T_\odot^4} = \frac{(R^2)(T_\odot/2)^4}{(R_\odot)^2 T_\odot^4} = \frac{R^2}{(R_\odot)^2 16} \\ \Rightarrow \frac{R^2}{(R_\odot)^2} &= 16 \times 400 \\ \Rightarrow \frac{R}{R_\odot} &= 80\end{aligned}$$

Nature of science

Reasoning about the universe

Over millennia, humans have mapped the planets and stars, recording their movements and relative brightness. By applying the simple idea of parallax, the change in position of a star in the sky at times six months apart, astronomers could begin to measure the distances to stars. Systematic measurement of the distances and the relative brightness of stars and galaxies, with increasingly sophisticated tools, has led to an understanding of a universe that is so large it is difficult to imagine.



? Test yourself

- Determine the distance to Procyon, which has a parallax of $0.285''$.
- The distance to Epsilon Eridani is 10.8 ly . Calculate its parallax angle.
- Betelgeuse has an angular diameter of $0.016''$ (that is, the angle subtended by the star's diameter at the eye of an observer) and a parallax of $0.0067''$.
 - Determine the distance of Betelgeuse from the Earth.
 - What is its radius in terms of the Sun's radius?
- A neutron star has an average density of about 10^{17} kg m^{-3} . Show that this is comparable to the density of an atomic nucleus.
- A sunspot near the centre of the Sun is found to subtend an angle of 4.0 arc seconds. Find the diameter of the sunspot.
- The resolution of the Hubble Space Telescope is about 0.05 arc seconds. Estimate the diameter of the smallest object on the Moon that can be resolved by the telescope. The Earth–moon distance is $D = 3.8 \times 10^8\text{ m}$.
- The Sun is at a distance of 28000 ly from the centre of the Milky Way and revolves around the galactic centre with a period of about 211 million years. Estimate from this information the orbital speed of the Sun and the total mass of the Milky Way.
- Describe, with the aid of a clear diagram, what is meant by the **parallax method** in astronomy.
 - Explain why the parallax method fails for stars that are very far away.
- The light from a star a distance of 70 ly away is received on Earth with an apparent brightness of $3.0 \times 10^{-8}\text{ W m}^{-2}$. Calculate the luminosity of the star.
- The luminosity of a star is $4.5 \times 10^{28}\text{ W}$ and its distance from the Earth is 88 ly . Calculate the apparent brightness of the star.
- The apparent brightness of a star is $8.4 \times 10^{-10}\text{ W m}^{-2}$ and its luminosity is $6.2 \times 10^{32}\text{ W}$. Calculate the distance to the star in light years.
- Two stars have the same size but one has a temperature that is four times larger.
 - Estimate how much more energy per second the hotter star radiates.
 - The apparent brightness of the two stars is the same; determine the ratio of the distance of the cooler star to that of the hotter star.
- Two stars are the same distance from the Earth and their apparent brightnesses are $9.0 \times 10^{-12}\text{ W m}^{-2}$ (star A) and $3.0 \times 10^{-13}\text{ W m}^{-2}$ (star B). Calculate the ratio of the luminosity of star A to that of star B.
- Take the surface temperature of our Sun to be 6000 K and its luminosity to be $3.9 \times 10^{26}\text{ W}$. Find, in terms of the solar radius, the radius of a star with:
 - temperature 4000 K and luminosity $5.2 \times 10^{28}\text{ W}$
 - temperature 9250 K and luminosity $4.7 \times 10^{27}\text{ W}$.
- Two stars have the same luminosity. Star A has a surface temperature of 5000 K and star B a surface temperature of 10000 K .
 - Suggest which is the larger star and by how much.
 - The apparent brightness of A is double that of B; calculate the ratio of the distance of A to that of B.
- Star A has apparent brightness $8.0 \times 10^{-13}\text{ W m}^{-2}$ and its distance is 120 ly . Star B has apparent brightness $2.0 \times 10^{-15}\text{ W m}^{-2}$ and its distance is 150 ly . The two stars have the same size. Calculate the ratio of the temperature of star A to that of star B.
- Calculate the apparent brightness of a star of luminosity $2.45 \times 10^{28}\text{ W}$ and a parallax of $0.034''$.

Learning objectives

- Describe the use of stellar spectra.
- Work with the HR diagram, including representation of stellar evolution.
- Apply the mass–luminosity relation.
- Describe Cepheid variables and their use as standard candles.
- Describe the nature of the stars in the main regions of the HR diagram.
- Understand the limits on mass for white dwarfs and neutron stars.

D2 Stellar characteristics and stellar evolution

This section deals with the lives of stars on the main sequence and their evolution away from it. We will see how **stellar spectra** may be used to determine the chemical composition of stars, and will study an important diagram called the Hertzsprung–Russell (HR) diagram. We will follow the **stellar evolution** on the HR diagram and meet important classes of stars such as Cepheid variables, **white dwarfs** and **red giants**.

D2.1 Stellar spectra

The energy radiated by a star is in the form of electromagnetic radiation and is distributed over an infinite range of wavelengths. A star is assumed to radiate like a black body. Figure D.9 shows black-body spectra at various temperatures.

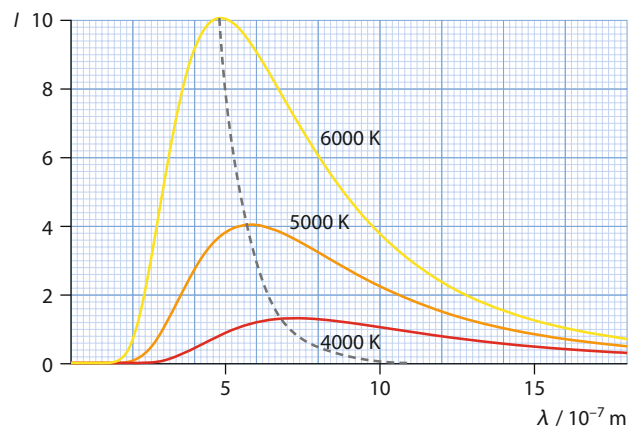


Figure D.9 Black-body radiation profiles at various temperatures. The broken lines show the variation with temperature of the peak intensity and the wavelength at which it occurs.

Much information can be determined about a star by examining its spectrum. The first piece of information is its surface temperature. Most of the energy is emitted around a wavelength called the peak wavelength. Calling this wavelength λ_0 , we see that the colour of the star is mainly determined by the colour corresponding to λ_0 .

The **Wien displacement law** relates the wavelength λ_0 to the surface temperature T :

$$\lambda_0 T = \text{constant} = 2.90 \times 10^{-3} \text{ K m}$$

This implies that the higher the temperature, the lower the wavelength at which most of the energy is radiated.



Worked examples

D.9 The Sun has an approximate black-body spectrum with most of its energy radiated at a wavelength of 5.0×10^{-7} m. Find the surface temperature of the Sun.

From Wien's law, $5.0 \times 10^{-7} \text{ m} \times T = 2.9 \times 10^{-3} \text{ K m}$; that is, $T = 5800 \text{ K}$.

D.10 The Sun (radius $R = 7.0 \times 10^8$ m) radiates a total power of 3.9×10^{26} W. Find its surface temperature.

From $L = \sigma AT^4$ and $A = 4\pi R^2$, we find

$$T = \left(\frac{L}{\sigma 4\pi R^2} \right)^{1/4} \approx 5800 \text{ K}$$

The surface temperature of a star is determined by measuring the wavelength at which most of its radiation is emitted (see Figure D.10).

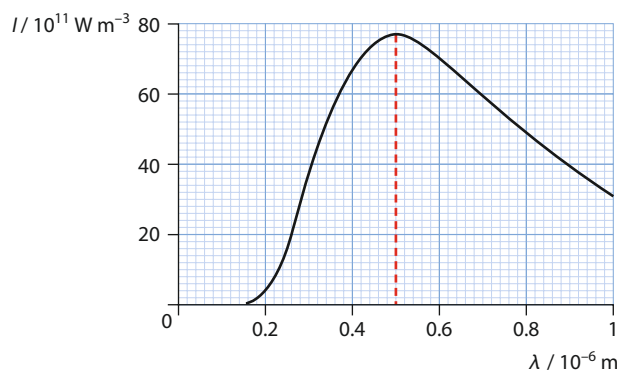


Figure D.10 The spectrum of this star shows a peak wavelength of 500 nm. Using Wien's law, we can determine its surface temperature.

The second important piece of information from a star's spectrum is its chemical composition. It is common to obtain an absorption spectrum in which dark lines are superimposed on a background of continuous colour (as in Figure D.11). Each dark line represents the absorption of light of a specific wavelength by a specific chemical element in the star's atmosphere.



Figure D.11 Absorption spectrum of a star, showing the absorption lines of hydrogen. A real spectrum would show very many dark lines corresponding to other elements as well.

It is known that most stars have essentially the same chemical composition, yet show different absorption spectra. The reason for this difference is that different stars have different temperatures. Consider two stars with the same content of hydrogen. One is hot, say at 25 000 K, and the other cool, say at 10 000 K. The hydrogen in the hot star is ionised, which means the electrons have left the hydrogen atoms. These atoms cannot absorb any light passing through them, since there are no bound electrons to absorb the photons and make transitions to higher-energy states. Thus, the hot star will not show any absorption lines at hydrogen

wavelengths. The cooler star, however, has many of its hydrogen atoms in the energy state $n=2$. Electrons in this state can absorb photons to make transitions to states such as $n=3$ and $n=4$, giving rise to characteristic hydrogen absorption lines. Similarly, an even cooler star – of temperature, say, 3000 K – will have most of the electrons in its hydrogen atoms in the ground state, so they can only absorb photons corresponding to ultraviolet wavelengths. These will not result in dark lines in an optical spectrum.

Stars are divided into seven **spectral classes** according to their colour (see Table D.3). As we have seen, colour is related to surface temperature. The spectral classes are called O, B, A, F, G, K and M (remembered as Oh Be A Fine Girl/Guy Kiss Me!).

It is known from spectral studies that hydrogen is the predominant element in normal **main-sequence** stars, making up upto 70% of their mass, followed by helium with at 28%; the rest is made up of heavier elements.

Spectral studies also give information on the star's velocity and rotation (through the Doppler shifting of spectral lines) and the star's magnetic field owing to the splitting of spectral lines in a magnetic field.

Spectral class	Colour	Temperature / K
O	Electric blue	25 000–50 000
B	Blue	12 000–25 000
A	White	7 500–12 000
F	Yellow–white	6 000–7 500
G	Yellow	4 500–6 000
K	Orange	3 000–4 500
M	Red	2 000–3 000

Table D.3 Colour and temperature characteristics of spectral classes.

D2.2 The Hertzsprung–Russell diagram

Astronomers realised early on that there was a correlation between the luminosity of a star and its surface temperature. In the early part of the twentieth century, the Danish astronomer Ejnar Hertzsprung and the American astronomer Henry Norris Russell independently pioneered plots of stellar luminosities. Such plots are now called **Hertzsprung–Russell (HR) diagrams**. In the HR diagram in Figure D.12, the vertical axis represents luminosity in units of the Sun's luminosity (that is, 1 on the vertical axis corresponds to the solar luminosity, $L_{\odot} = 3.9 \times 10^{26}$ W). The horizontal axis shows the surface temperature of the star (in kelvin). The temperature decreases to the right.

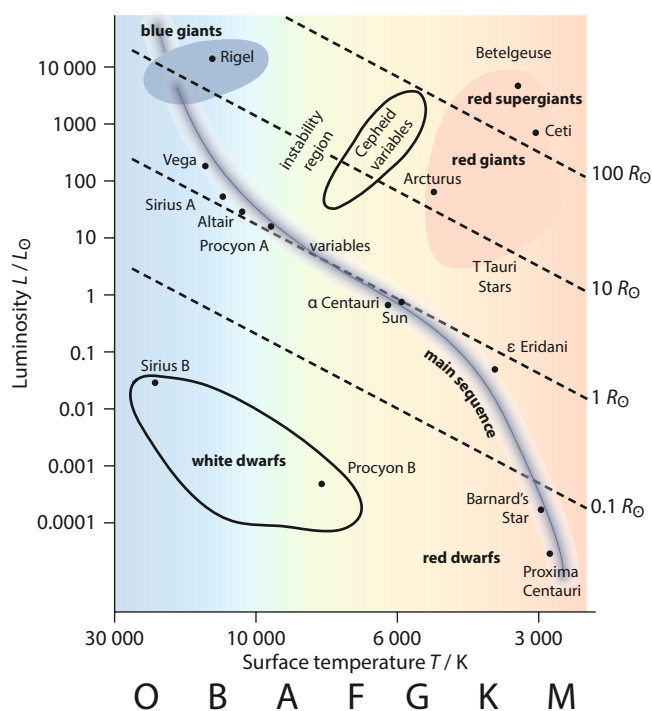


Figure D.12 A Hertzsprung–Russell diagram. Surface temperature increases to the left. Note that the scales are not linear.

Exam tip

It is very important that you clearly understand the HR diagram.



Also shown is the spectral class, which is an alternative way to label the horizontal axis. The luminosity in this diagram varies from 10^{-4} to 10^4 , a full eight orders of magnitude, whereas the temperature varies from 3000 K to 30 000 K. For this reason, the scales on each axis are not linear.

The slanted dotted lines represent stars with the same radius, so our Sun and Procyon have about the same radius. The symbol R_{\odot} stands for the radius of the Sun.

As more and more stars were placed on the HR diagram, it became clear that a pattern was emerging. The stars were not randomly distributed on the diagram. Three clear features emerge:

- Most stars fall on a strip extending diagonally across the diagram from top left to bottom right. This is called the **main sequence**.
- Some large stars, reddish in colour, occupy the top right. These are the **red giants** (large and cool). Above these are the **red supergiants** (very large and cool).
- The bottom left is a region of small stars known as **white dwarfs** (small and hot).

As we will see in Section D2.3, the higher the luminosity of a main-sequence star, the higher its mass. So as we move along the main sequence towards hotter stars, the masses of the stars increase. Thus, the right end of the main sequence is occupied by **red dwarfs** and the left by **blue giants**.

Note that, once we know the temperature of a star (for example, through its spectrum), the HR diagram can tell us its luminosity with an acceptable degree of accuracy, provided it is a main-sequence star.

D2.3 Main-sequence stars

Our Sun is a typical member of the main sequence. It has a mass of 2×10^{30} kg, a radius of 7×10^8 m and an average density of 1.4×10^3 kg m⁻³, and it radiates at a rate of 3.9×10^{26} W. What distinguishes different main-sequence stars is their mass (see Figure D.13). Main-sequence stars produce enough energy in their core, from the nuclear fusion of hydrogen into helium, to exactly counterbalance the tendency of the star to collapse under its own weight. The common characteristic of all main-sequence stars is the fusion of hydrogen into helium.

D2.4 Red giants and red supergiants

Red giants are very large, cool stars with a reddish appearance. The luminosity of red giants is considerably greater than that of main-sequence stars of the same temperature. Treating them as black bodies radiating according to the Stefan–Boltzmann law means that a luminosity which is 10^3 times greater than that of our Sun corresponds to a surface area which is 10^3 times that of the Sun, and thus a radius about 30 times greater. The mass of a red giant can be as much as 100 times the mass of our Sun, but their huge size also implies small densities. A red giant will have a central hot core surrounded by an enormous envelope of extremely tenuous gas.

Red supergiants are even larger. Extreme examples include stars with radii that are 1500 times that of our Sun and luminosities of 5×10^5 solar luminosities.

Exam tip

Main-sequence stars: fuse hydrogen to form helium,
 Red giants: bright, large, cool, reddish, tenuous.
 Red supergiants: even larger and brighter than red giants,
 White dwarfs: dim, small, hot, whitish, dense.

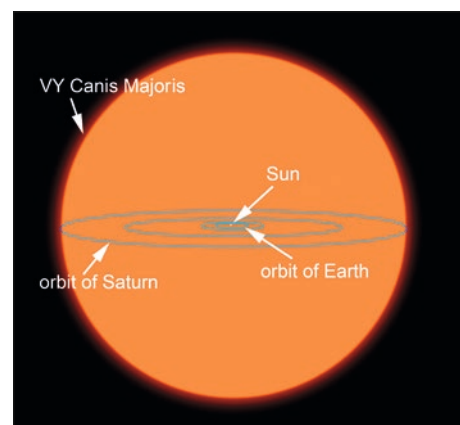


Figure D.13 Artist's impression of a comparison between our Sun and the red supergiant VY Canis Majoris.

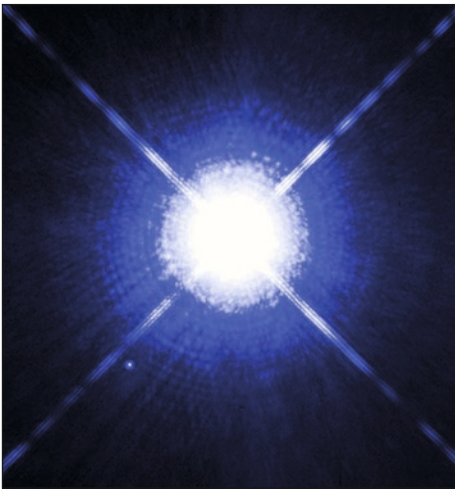


Figure D.14 Sirius A is the bright star in the middle of the photograph. Its white dwarf companion, Sirius B, is the tiny speck of light at the lower left. The rings and spikes are artefacts of the telescope's imaging system. The photograph has been overexposed so that the faint Sirius B can be seen.

D2.5 White dwarfs

White dwarf stars are common but their faintness makes them hard to detect. A well-known white dwarf is Sirius B, the second star in a binary star system (double star) whose other member, Sirius A, is the brightest star in the evening sky (Figure D.14).

Sirius A and Sirius B have about the same surface temperature (about 10 000 K) but the luminosity of Sirius B is about 10 000 times smaller. This means that it has a radius that is 100 times smaller than that of Sirius A. Here is a star with a mass roughly that of the Sun with a size similar to that of the Earth. This means that its density is about 10^6 times the density of the Earth!

Worked example

D.11 A main-sequence star emits most of its energy at a wavelength of 2.4×10^{-7} m. Its apparent brightness is measured to be $4.3 \times 10^{-9} \text{ W m}^{-2}$. Estimate the distance of the star.

From Wien's law, we find the temperature of the star to be given by

$$\begin{aligned}\lambda_0 T &= 2.9 \times 10^{-3} \text{ K m} \\ \Rightarrow T &= \frac{2.9 \times 10^{-3}}{2.4 \times 10^{-7}} \text{ K} \\ &= 12\,000 \text{ K}\end{aligned}$$

From the HR diagram in Figure D.12, we see that such a temperature corresponds to a luminosity about 100 times that of the Sun: that is, $L = 3.9 \times 10^{28} \text{ W}$. Thus,

$$\begin{aligned}d &= \sqrt{\frac{L}{4\pi b}} = \sqrt{\frac{3.9 \times 10^{28}}{4\pi \times 4.3 \times 10^{-9}}} \text{ m} \\ &= 8.5 \times 10^{17} \text{ m} \approx 90 \text{ ly} \approx 28 \text{ pc}\end{aligned}$$

D2.6 The mass–luminosity relation

For stars on the main sequence, there exists a relation between the mass and the luminosity of the star. The **mass–luminosity relation** states that

$$L \propto M^{3.5}$$

This relation comes from application of the laws of nuclear physics to stars. Main-sequence stars in the upper left-hand corner of the HR diagram have a very high luminosity and therefore are very massive.

Exam tip

The mass–luminosity relation can only be used for main-sequence stars.

Worked example

D.12 Use the HR diagram and the mass–luminosity relation to estimate the ratio of the density of Altair to that of the Sun.

The two stars have the same radius and hence the same volume. The luminosity of Altair is about 10 times that of the Sun. From

$$\frac{L}{L_{\odot}} = \left(\frac{M}{M_{\odot}}\right)^{3.5}$$

we get

$$10 = \left(\frac{M}{M_{\odot}}\right)^{3.5} \Rightarrow \frac{M}{M_{\odot}} = 10^{1/3.5} \approx 1.9$$

Hence the ratio of densities is also about 1.9, since the volumes are the same.

D2.7 Cepheid stars

Cepheid variable stars are stars whose luminosity is not constant in time but varies **periodically** from a minimum to a maximum, the periods being typically from a couple of days to a couple of months. The brightness of the star increases sharply and then fades off more gradually, as shown in the **light curve** of a Cepheid in Figure D.15.

The mechanism for the periodic variation of the luminosity of Cepheid stars is the following. As radiation rushes outwards it ionises helium atoms in the atmosphere of the star. The freed electrons, through collisions, heat up the star's atmosphere. This increases the pressure, which forces the outer layers of the star to expand. When most of the helium is ionised, radiation now manages to leave the star, and the star cools down and begins to contract under its own weight. This makes helium nuclei recombine with electrons, and so the cycle repeats as helium can again be ionised. The star is brightest when the surface is expanding outwards at maximum speed.

Cepheids occupy a strip between the main sequence and the red giants on an HR diagram.

At the beginning of the 20th century, astronomer Henrietta Leavitt discovered a remarkably precise relationship between the average luminosity of Cepheids and their period. The longer the period, the larger the luminosity (see Figure D.16). This makes Cepheid stars **standard candles** – that is, stars of a known luminosity, obtained by measuring their period.

Exam tip

The reason for a Cepheid star's periodic variation in luminosity is the periodic expansion and contraction of the outer layers of the star.

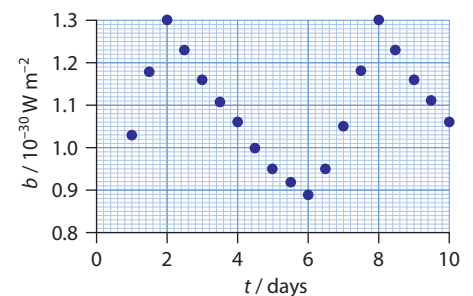


Figure D.15 The apparent brightness of a Cepheid star varies periodically with time.

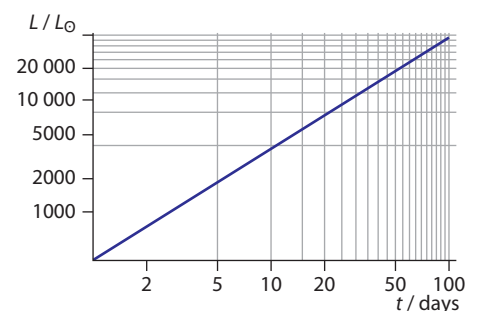


Figure D.16 The relationship between peak luminosity and period for Cepheid stars.

Worked example

D.13 Estimate the distance of the Cepheid whose light curve is shown in Figure D.15.

The period is 6 days. From Figure D.16, this corresponds to a luminosity of about 2000 solar luminosities, or about $L = 7.2 \times 10^{29}$ W. The average apparent brightness is $b = 1.1 \times 10^{-10}$ W m⁻². Therefore

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

$$\Rightarrow d = \sqrt{\frac{L}{4\pi b}} = \sqrt{\frac{7.2 \times 10^{29}}{4\pi \times 1.1 \times 10^{-10}}} \text{ m}$$

$$= 2.28 \times 10^{19} \text{ m} \approx 2400 \text{ ly} \approx 740 \text{ pc}$$

Thus, one can determine the distance to the galaxy in which a Cepheid is assumed to be. The Cepheid method can be used to find distances up to a few megaparsecs.

D2.8 Stellar evolution: the Chandrasekhar and Oppenheimer–Volkoff limits

Stars are formed out of contracting gases and dust in the **interstellar medium**, which has hydrogen as its main constituent. Initially the star has a low surface temperature and so its position is somewhere to the right of the main sequence on the HR diagram. As the star contracts under its own weight, gravitational potential energy is converted into thermal energy and the star heats up; it begins to move towards the main sequence. The time taken to reach the main sequence depends on the mass of the star; heavier stars take less time. Our Sun, an average star, has taken about 20 to 30 million years (see Figure D.17).

As a star is compressed more and more (under the action of gravity), its temperature rises and so does its pressure. Eventually, the temperature in the core reaches 5×10^6 to 10^7 K and nuclear fusion reactions commence, resulting in the release of enormous amounts of energy. The energy released can account for the sustained luminosity of stars such as our Sun, for example, over the 4–5 billion years of its life so far. Thus, nuclear fusion provides the energy that is needed to keep the star hot, so that its pressure is high enough to oppose further contraction, and at the same time to provide the energy that the star is radiating into space.

On the main sequence, the main nuclear fusion reactions are those of the proton–proton cycle (Section D1.2), in which the net effect is to turn four hydrogen nuclei into one helium-4 nucleus.

When about 12% of the hydrogen in the star has been used up in nuclear fusion, a series of instabilities develops in the star, upsetting the delicate balance between radiation pressure and gravitational pressure. The star will then begin to move away from the main sequence. What happens next is determined mainly by the mass of the star. Other types of nuclear fusion reactions will take place (see Section D4) and the star will change in size and surface temperature (and hence colour).

The changes that take place can be shown as paths on the HR diagram. We may distinguish two essentially different paths, the first for what we will call **low-mass** stars, with a mass less than about eight

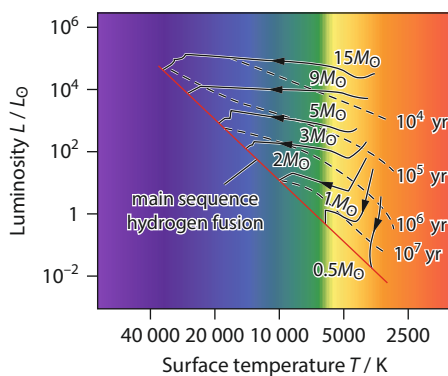


Figure D.17 Evolutionary tracks of protostars as they approach the main sequence. M_{\odot} stands for one solar mass.

The mass of a star is the main factor that determines its evolution off the main sequence.

solar masses, and the second for stars with a higher mass. (In reality the situation is more complex, but this simple distinction is sufficient for the purposes of this course.)

In low-mass stars, helium collects in the core of the star, surrounded by a thin shell of hydrogen and a bigger hydrogen envelope (Figure D.18).

Only hydrogen in the thin inner shell undergoes nuclear fusion to helium. The temperature and pressure of the helium build up and eventually helium itself begins to fuse (this is called the 'helium flash'), with helium in a thin inner shell producing carbon in the core. In the core, some carbon nuclei fuse with helium to form oxygen. Oxygen is the heaviest element that can be produced in low-mass stars; the temperature never rises enough for production of heavier elements. The hydrogen in the thin shell is still fusing, so the star now has nuclear fusion in two shells, the H and He shells. The huge release of energy blows away the outer layers of the star in an explosion called a **planetary nebula**; mass is thrown into space, leaving behind the carbon core (and some oxygen). This evolution may be shown on an HR diagram (Figure D.19). We will return to the processes in the core in Section D4.2.

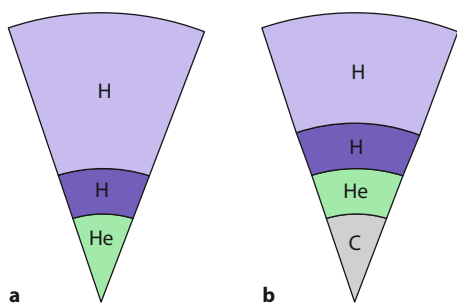


Figure D.18 a The structure of a low-mass star after it leaves the main sequence. b After helium begins to fuse, carbon collects in the core.

The path takes the star off the main sequence and into the red giant region. The star gets bigger and cooler on the surface and hence becomes red in colour. The time taken to leave the main sequence and reach the planetary nebula stage is short compared with the time spent on the main sequence: it takes from a few tens to a few hundreds of million of years. The path then takes the star to the white dwarf region. The star is now a stable but dead star (Figure D.20). No nuclear reactions take place in the core.

The conditions in the core mean that the electrons behave as a gas, and the pressure they generate is what keeps the core from collapsing further under its weight. This pressure is called **electron degeneracy pressure** and is the result of a quantum mechanical effect, referred to as the Pauli Exclusion Principle, which states that no two electrons may occupy the same quantum state.

The core has now become a **white dwarf** star. Now exposed, and with no further energy source, the star is doomed to cool down to practically zero temperature and will then become a **black dwarf**.

Electron degeneracy pressure prevents the further collapse of the core and – provided the mass of the core is less than about 1.4 solar masses – the star will become a stable white dwarf. This important number is known in astrophysics as the **Chandrasekhar limit**.

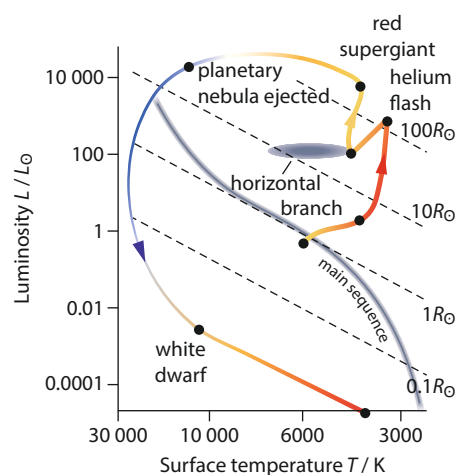


Figure D.19 Evolutionary path of a low-mass star. This is the path of a star of one solar mass that ends up as a white dwarf, which continues to cool down, moving the star ever more to the right on the HR diagram.



Figure D.20 The Helix, a planetary nebula. The star that produced this nebula can be seen at its exact centre.

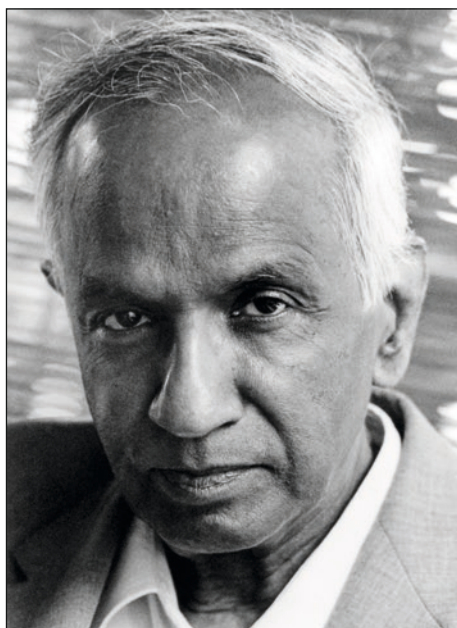


Figure D.21 Subrahmanyan Chandrasekhar (1910–1995).

Exam tip

Fusion ends with the production of iron.

Exam tip

This summary and the paths on the HR diagram are the bare minimum you should know for an exam.

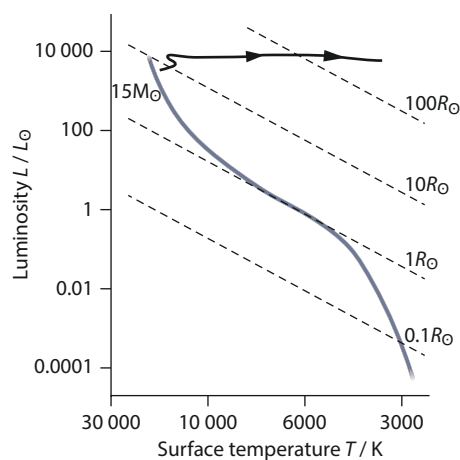


Figure D.22 Evolutionary path of a star of 15 solar masses. It becomes a red supergiant that explodes in a supernova. After the supernova, the star becomes a neutron star, whose luminosity is too small to be plotted on the HR diagram.

The limit is named after the astrophysicist Subrahmanyan Chandrasekhar (Figure D.21), who discovered it in the 1930s.

We now look at the evolution of stars whose mass is greater than about eight solar masses. The process begins much the same way as it did for low-mass stars, but differences begin to show when carbon fuses with helium in the core to form oxygen. If the mass of the star is large enough, the pressure caused by gravity is enough to raise the temperature sufficiently to allow the formation of ever-heavier elements: neon, more oxygen, magnesium and then silicon; eventually iron is produced in the most massive stars, and that is where the process stops, since iron is near the peak of the binding-energy curve. It would require additional energy to be supplied for iron to fuse.

The star moves off the main sequence and into the red supergiant area (Figure D.22). As the path moves to the right, ever-heavier elements are produced. The star is very hot in the core. Photons have enough energy at these temperatures to split nuclei apart; in about one second (!) millions of years, worth of nuclear fusion is undone. Nuclei are in turn ripped apart into individual protons and neutrons, so that in a very short time the star is composed mainly of protons, electrons, neutrons and photons.

Because of the high densities involved, the electrons are forced into the protons, turning them into neutrons and producing neutrinos that escape from the star ($e^- + p \rightarrow n + \nu_e$). The star's core is now made up almost entirely of neutrons, and is still contracting rapidly. The Pauli Exclusion Principle may now be applied to the neutrons: if they get too close to one another, a pressure develops to prevent them from getting any closer. But they have already done so, and so the entire core now rebounds to a larger equilibrium size. This rebound is catastrophic for the star, creating an enormous shock wave travelling outwards that tears the outer layers of the star apart. The resulting explosion, called a **supernova**, is much more violent than a planetary nebula. The energy loss from this explosion leads to a drastic drop in the temperature of the star, and it begins to collapse.

The core that is left behind, which is more massive than the Chandrasekhar limit, will most likely become a **neutron star**. Neutron pressure keeps such a star stable, provided the mass of the core is not more than about 2–3 solar masses – the **Oppenheimer–Volkoff limit**. If its mass higher than this, it may collapse further and become a black hole.

Table D.4 shows the temperatures at which various elements participate in fusion reactions.

Element	$T / 10^6 \text{ K}$	Where
Hydrogen	1–20	Main sequence
Helium	100	Red giant
Carbon	500–800	Supergiant
Oxygen	1000	Supergiant

Table D.4 Temperatures at which various elements participate in fusion reactions.



To summarise:

If the mass of the core of a star is less than the Chandrasekhar limit of about 1.4 solar masses, it will become a stable white dwarf, in which electron pressure keeps the star from collapsing further.

If the core is more massive than the Chandrasekhar limit but less than the Oppenheimer–Volkoff limit of about 2–3 solar masses, the core will collapse further until electrons are driven into protons, forming neutrons. Neutron pressure now keeps the star from collapsing further, and the star becomes a neutron star.

If the Oppenheimer–Volkoff limit is exceeded, the star will become a black hole.

Initial mass of star (in terms of solar masses)	Outcome
0.08–0.25	White dwarf with helium core
0.25–8	White dwarf with carbon core
8–12	White dwarf with oxygen/neon/magnesium core
12–40	Neutron star
>40	Black hole

Table D.5 The final fate of stars with various initial masses.

Table D.5 shows the final end products of evolution for different initial stellar masses. Figure D.23 is a schematic summary of the life history of a star.

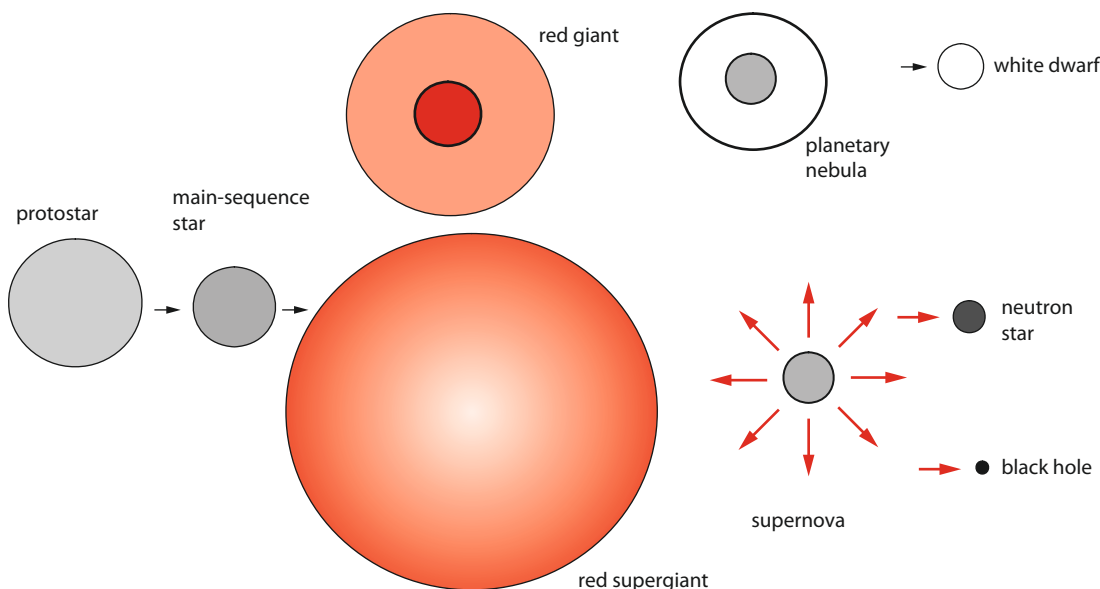


Figure D.23 The birth and death of a star. The star begins as a protostar, evolves to the main sequence and then becomes a red giant or supergiant. After a planetary nebula or supernova explosion, the core of the star develops into one of the three final stages of stellar evolution: white dwarf, neutron star or black hole.

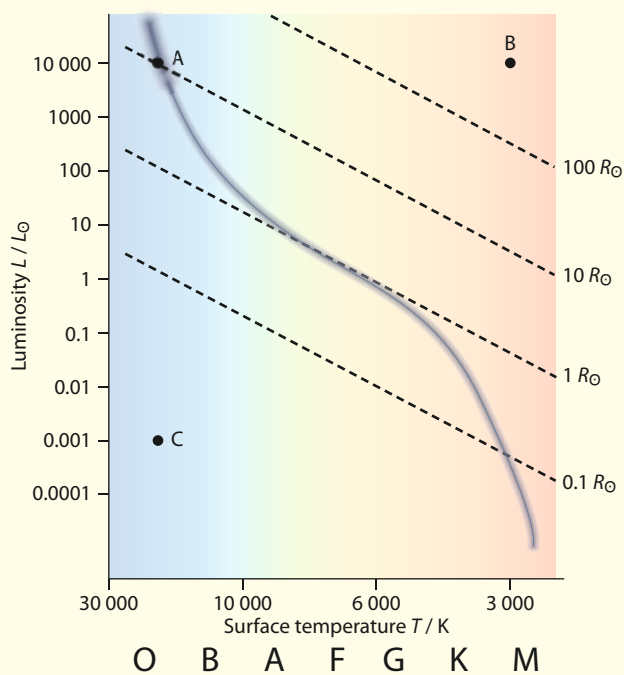
Nature of science

Evidence from starlight

The light from a star is the best source of information about it. The distribution of frequencies tells us its surface temperature, and the actual frequencies present tell us its composition, as each element has a characteristic spectrum. The luminosity and temperature of a star are related, and together give us information about the evolution of stars of different masses. Using this evidence, Chandrasekhar predicted a limit to the mass of a star that would become a white dwarf, while Oppenheimer and Volkoff predicted the mass above which it would become a black hole. The development of theories of stellar evolution illustrates how, starting from simple observations of the natural world, science can build up a detailed picture of how the universe works. Further observations are then needed to confirm or reject hypotheses.

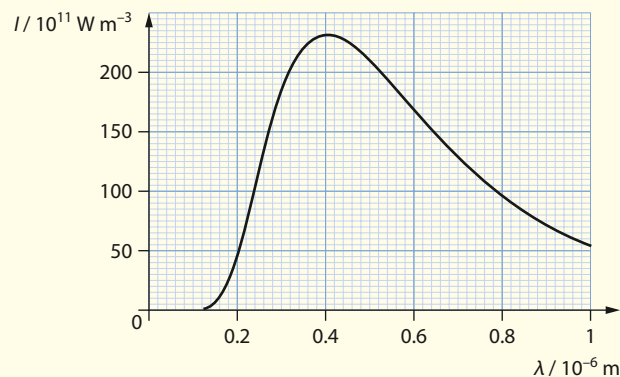
? Test yourself

- 18 Describe how a stellar absorption spectrum is formed.
- 19 Describe how the chemical composition of a star may be determined.
- 20 Describe how the colour of the light from a star can be used to determine its surface temperature.
- 21 Stars A and B emit most of their light at wavelengths of 650 nm and 480 nm, respectively. Star A has twice the radius of star B. Find the ratio of the luminosities of the stars.
- 22 **a** State what is meant by a Hertzsprung–Russell (HR) diagram.
b Describe the main features of the HR diagram.
c The luminosity of the Sun is $3.9 \times 10^{26} \text{ W}$ and its radius is $7.0 \times 10^8 \text{ m}$. For star A in the HR diagram below **calculate**
i the temperature
ii the density in terms of the Sun's density.
d For stars B and C calculate the radius in terms of the Sun's radius.



- 23 A main-sequence star emits most of its energy at a wavelength of $2.42 \times 10^{-7} \text{ m}$. Its apparent brightness is measured to be $8.56 \times 10^{-12} \text{ W m}^{-2}$. Estimate its distance using the HR diagram in question 22.

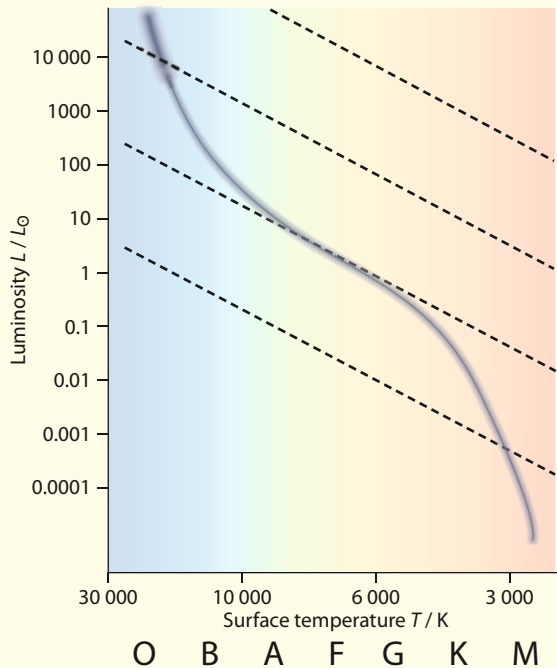
- 24 A main-sequence star is 15 times more massive than our Sun. Calculate the luminosity of this star in terms of the solar luminosity.
- 25 **a** The luminosity of a main-sequence star is 4500 times greater than the luminosity of our Sun. Estimate the mass of this star in terms of the solar mass.
b A star has a mass of 12 solar masses and a luminosity of 3200 solar luminosities. Determine whether this could be a main-sequence star.
- 26 Describe the mechanism by which the luminosity of Cepheid stars varies.
- 27 Using Figure D.16, calculate the distance of a Cepheid variable star whose period is 10 days and whose average apparent brightness is $3.45 \times 10^{-14} \text{ W m}^{-2}$.
- 28 **a** Find the temperature of a star whose spectrum is shown below.



- b** Assuming this is a main-sequence star, estimate its luminosity using the HR diagram in question 22.
- 29 Estimate the temperature of the universe when the peak wavelength of the radiation in the universe was $7.0 \times 10^{-7} \text{ m}$.
- 30 A neutron star has a radius of 30 km and makes 500 revolutions per second.
a Calculate the speed of a point on its equator.
b Determine what fraction of the speed of light this is.
- 31 Describe the formation of a red giant star.
- 32 **a** Describe what is meant by a **planetary nebula**.
b Suggest why most photographs show planetary nebulae as rings: doesn't the gas surround the core in all directions?



- 33 Assume that no stars of mass greater than about two solar masses could form anywhere. Would life as we know it on Earth be possible?
- 34 Describe the evolution of a main-sequence star of mass:
- 2 solar masses
 - 20 solar masses.
 - Show the evolutionary paths of these stars on a copy of the HR diagram below.



- Describe the formation of a white dwarf star.
 - List two properties of a white dwarf.
 - Describe the mechanism which prevents a white dwarf from collapsing under the action of gravity.
- 36 Describe **two** differences between a main-sequence star and a white dwarf.
- 37 A white dwarf, of mass half that of the Sun and radius equal to one Earth radius, is formed. Estimate its density.

- 38 Describe **two** differences between a main-sequence star and a neutron star.
- 39
- Describe the formation of a neutron star.
 - List two properties of a neutron star.
 - Describe the mechanism which prevents a neutron star from collapsing under the action of gravity.
- 40 Describe your understanding of the **Chandrasekhar limit**.
- 41 Describe your understanding of the **Oppenheimer-Volkoff limit**.
- 42 Assume that the material of a main-sequence star obeys the ideal gas law, $PV = NkT$. The volume of the star is proportional to the cube of its radius R , and N is proportional to the mass M of the star.
- Show that $PR^3 \propto MT$.
The star is in equilibrium under the action of its own gravity, which tends to collapse it, and the pressure created by the outflow of energy from its interior, which tends to expand it. It can be shown that this equilibrium results in the condition $P \propto \frac{M^2}{R^2}$. (**Can you see how?**)
 - Combine these two proportionalities to show that $P \propto \frac{M}{R}$. Use this result to explain that, as a star shrinks, its temperature goes up.
 - Conclude this rough analysis by showing that the luminosity of main-sequence stars of the same density is given by $L \propto M^{3.3}$.

Learning objectives

- Understand Hubble's law.
- Understand the scale factor and red-shift.
- Understand the cosmic microwave background radiation.
- Understand the accelerating universe and red-shift.

D3 Cosmology

This section deals with three dramatic discoveries in cosmology: the discovery of the expansion of the universe by Hubble, the discovery of the **cosmic background radiation** by Penzias and Wilson, and finally the discovery of the accelerated rate of expansion by Perlmutter, Schmidt and Riess.

D3.1 Hubble's law: the expanding universe

Early in the 20th century, studies of galaxies revealed red-shifted absorption lines. Application of the standard Doppler effect indicated that these galaxies were moving **away from us**. By 1925, 45 galaxies had been studied, and all but the closest ones appeared to be moving away at enormous speeds.



Physics in distant galaxies is the same as that on Earth

How do we know the wavelength of light emitted by distant galaxies? Light emitted from galaxies comes from atomic transitions in the hot gas in the interior of the galaxies, which is mostly hydrogen. Galaxies are surrounded by cooler gas and thus light travelling through is absorbed at specific wavelengths, showing a characteristic absorption spectrum. The wavelengths corresponding to the dark lines are well known from experiments on Earth.

The velocity of recession is found by an application of the Doppler effect to light. Light from galaxies arrives on Earth red-shifted. This means that the wavelength of the light measured upon arrival is longer than the wavelength at emission. According to the Doppler effect, this implies that the source of the light – the galaxy – is moving away from observers on Earth.

The Doppler effect *may* be used to describe the red-shift in the light from distant galaxies. However, the red-shift is a consequence of the expanding universe in the sense that the space between galaxies is stretching out (expands) and this gives the illusion of galaxies moving away from each other; see Section **D3.2**.

If λ_0 is the wavelength of a spectral line and λ is the (longer) wavelength received on Earth, the red-shift z of the galaxy is defined as

$$z = \frac{\lambda - \lambda_0}{\lambda_0}$$

If the speed v of the receding galaxy is small compared with the speed of light c , then the Doppler formula is $z = \frac{v}{c}$, which shows that the red-shift is indeed directly proportional to the receding galaxy's speed (more correctly, the component of its velocity along the line of sight).

In 1925, Edwin Hubble began a study to measure the distance to the galaxies for which the velocities of recession had been determined. In

Exam tip

Notice that $z = \frac{\lambda - \lambda_0}{\lambda_0}$ can also

be written as $z = \frac{\lambda}{\lambda_0} - 1$.

The proper interpretation of the red-shift is not through the Doppler effect (even though we use the Doppler formula) but the stretching of space in between galaxies as the universe expands.



1929, Hubble announced that distant galaxies move away from us with speeds that are proportional to their distance (Figure D.24).

Hubble studied a large number of galaxies and found that the more distant the galaxy, the faster it moves away from us. This is **Hubble's law**, which states that the velocity of recession is directly proportional to the distance, or

$$v = H_0 d$$

where d is the distance between the Earth and the galaxy and v is the galaxy's velocity of recession. The constant of proportionality, H_0 , is the slope of the graph and is known as the **Hubble constant**.

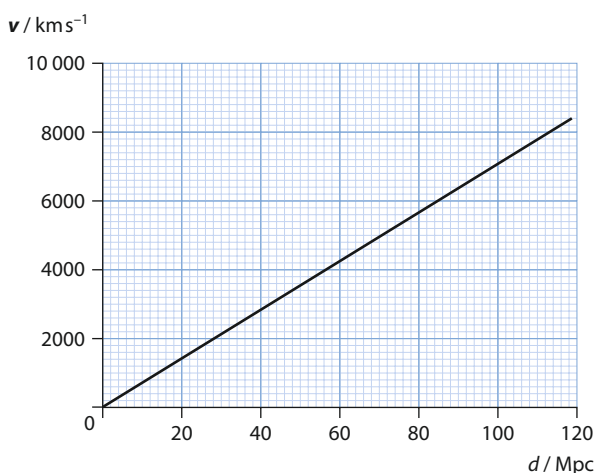


Figure D.24 Hubble discovered that the velocity of recession of galaxies is proportional to their distance from us.

Using $z = \frac{v}{c}$, we can rewrite Hubble's law as

$$z = \frac{H_0 d}{c} \Rightarrow d = \frac{cz}{H_0}$$

This formula relates distance to red-shift. It is an approximate relation, valid only for values of the red-shift z up to about 0.2.

There has been considerable debate as to the value of the Hubble constant. The most recent value, provided by the ESA's Planck satellite observatory data, is $H_0 = 67.80 \pm 0.77 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

Worked example

D.14 A hydrogen line has a wavelength of 434 nm. When received from a distant galaxy, this line is measured on Earth to be at 486 nm. Calculate the speed of recession of this galaxy.

The red-shift is $\frac{486 - 434}{434} = 0.12$, so $v = 0.12c = 3.6 \times 10^7 \text{ ms}^{-1}$.

Hubble's discovery implies that, in the past, the distances between galaxies were smaller and, moreover, that at a specific time in the past the entire universe had the size of a point. This specific time is taken to be the beginning of the universe, and leads to the model of the universe known as the **Big Bang model**, to be discussed in Section **D3.3**. Not only time, but also the space in which the matter and energy of the universe reside were created at that moment. As the space expanded, the distance between clumps of matter increased, leading to the receding galaxies that Hubble observed.

Hubble's law does not imply that the Earth is at the centre of the universe, even though the observation of galaxies moving away from us might lead us to believe so. An observer on a different star in a different galaxy would reach the same (erroneous) conclusion about their location.

D3.2 The cosmic scale factor R and red-shift

The expansion of the universe can be described in terms of a scale factor, R . To understand what this is, consider two points with coordinates 1 and 3 on a number line (Figure **D.25a**). In ordinary geometry we would have no problem saying that the distance between the two points is the difference in their coordinates, $3 - 1 = 2$ units. Let us call this difference in coordinates Δx . If space expands, however, after some time the diagram would look like Figure **D.25b**. The distance between the points has increased but the difference in their coordinates has remained the same. So this difference does not give the actual physical distance between the points.

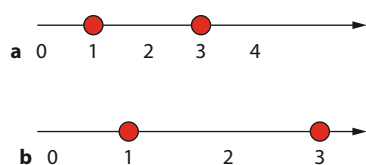


Figure D.25 As the space between two points stretches, the physical distance between them increases, even though the coordinates of the points do not change.

The meaning of the scale factor R is that multiplying the difference in coordinates Δx by R gives the physical distance d between the points:

$$d = R \Delta x$$

The scale factor may depend on time. The function $R(t)$ is called the **scale factor** of the universe and is of basic importance to cosmology. It is sometimes referred to (very loosely) as the **radius of the universe**.

This gives a new and completely different interpretation of red-shift. Suppose that, when a photon of cosmic microwave background (CMB) radiation was emitted in the very distant past, its wavelength was λ_0 . Let Δx stand for the difference in the coordinates of two consecutive wave crests. Then

$$\lambda_0 = R_0 \Delta x$$

where R_0 is the value of the scale factor at the time of emission. If this same wavelength is observed now, its value will be

$$\lambda = R \Delta x$$

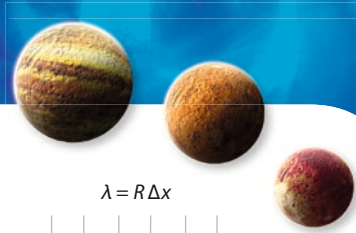
where R is the value of the scale factor at the present time. We deduce that

$$\frac{\lambda}{\lambda_0} = \frac{R}{R_0}$$

Thus the explanation for the observed red-shift in the light received from distant galaxies is not that the galaxies are really moving but that the space in between us and the galaxies is stretching (i.e. expanding).

Exam tip

SL students should know the result $z = \frac{R}{R_0} - 1$ but will not be examined on its derivation. HL students must know its derivation.



Of course, the stretching of space does give the illusion of motion, which is why applying the Doppler formula also gives the red-shift. But it must be stressed that the real explanation of the red-shift is the expansion of space and not the Doppler effect (see Figure D.26).

So the red-shift formula $z = \frac{\lambda - \lambda_0}{\lambda_0}$ becomes $z = \frac{R}{R_0} - 1$.

One of the great problems in cosmology is to determine how the scale factor depends on time. We will look at this problem in Section D5.2.

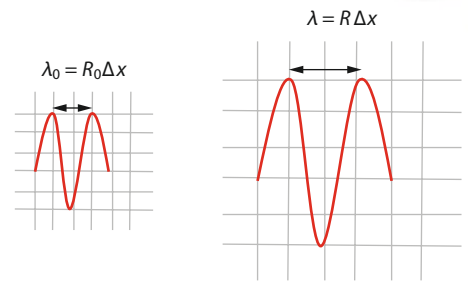


Figure D.26 As the universe expands, the wavelength of a photon emitted in the distant past increases when measured at the present time.

Worked examples

D.15 The peak wavelength of the CMB radiation at present is about 1 mm. In the past there was a time when the peak wavelength corresponded to blue light (400 nm). Estimate the size of the universe then, compared with its present size.

From $\frac{\lambda}{\lambda_0} = \frac{R}{R_0}$ we find

$$\frac{R}{R_0} = \frac{1 \times 10^{-3}}{4 \times 10^{-7}} \approx 2 \times 10^3$$

So when the universe was bathed in blue light it was smaller by a factor of about 2000.

D.16 Determine the size of the universe relative to its present size at the time when the light of Worked example D.14 was emitted.

The red-shift was calculated to be $z = 0.12$, so from $z = \frac{R}{R_0} - 1$ we find

$$0.12 = \frac{R}{R_0} - 1 \quad \Rightarrow \quad \frac{R}{R_0} = 1.12$$

The universe was 1.12 times smaller or $\frac{100\%}{1.12} = 89\%$ of its present size when the light was emitted.

D3.3 The Hot Big Bang model: the creation of space and time

The discovery of the expanding universe by Hubble implies a definite beginning, some 13.8 billion years ago. The size of the universe at that time was infinitesimally small and the temperature was enormous. Time, space, energy and mass were created at that instant. It is estimated that just 10^{-44} s after the beginning (the closest to $t=0$ about which something remotely reliable may be said), the temperature was of order 10^{32} K. These conditions create the picture of a gigantic explosion at $t=0$, which set matter moving outwards. Billions of years later, we see the remnants of this explosion in the receding motion of the distant galaxies. This is known as the **Hot Big Bang** scenario in cosmology.

The Big Bang was not an explosion that took place at a specific time in the past somewhere in the universe. At the time of the Big Bang, the



Why is the night sky dark?

The astronomers de Cheseaux and Olbers asked the very simple question of why the night sky is dark. Their argument, based on the prevailing static and infinite cosmology of the period, led to a night sky that would be uniformly bright! In its simplest form, the argument says that no matter where you look you will end up with a star. Hence the night sky should be uniformly bright, which it is not. This is Olbers' Paradox.

space in which the matter of the universe resides was created as well. Thus, the Big Bang happened about 13.8 billion years ago everywhere in the universe (the universe then being a point).

It is important to understand that the universe is not expanding into empty space. The expansion of the universe is not supposed to be like an expanding cloud of smoke that fills more and more volume in a room. The galaxies that are moving away from us are not moving into another, previously unoccupied, part of the universe. *Space is being stretched* in between the galaxies and so the distance between them is increasing, creating the illusion of motion of one galaxy relative to another.

There is plenty of experimental evidence in support of the Big Bang model. The first is the observation of an expanding universe that we have already talked about. The next piece of evidence is the cosmic microwave background radiation, to be discussed in Section D3.4.

If we assume that the expansion of the universe has been constant up to now, then $\frac{1}{H_0}$ gives an upper bound on the **age of the universe**. This is only an upper bound, since the fact that expansion rate was faster at the beginning implies a younger universe. The time $\frac{1}{H_0}$, known as the **Hubble time**, is about 14 billion years. The universe cannot be older than that. A more detailed argument that leads to this conclusion is as follows. Imagine a galaxy which is now at a distance d from us. Its velocity is thus $v = H_0 d$. In the beginning the galaxy and the Earth were at zero separation from each other. If the present separation of d was thus covered at the same constant velocity $H_0 d$, the time T taken to achieve this separation must be given by $H_0 d = \frac{d}{T}$, that is, $T = \frac{1}{H_0}$. T is thus a measure of the age of the universe.

The numerical value of the Hubble time with $H_0 = 67.80 \times 10^3 \text{ ms}^{-1} \text{ Mpc}^{-1}$ is

$$\begin{aligned} T_H &= \frac{1}{H} \\ &= \frac{1}{67.80 \times 10^3 \text{ ms}^{-1} \text{ Mpc}^{-1}} = \frac{1}{67.80 \times 10^3 \text{ ms}^{-1}} \times 10^6 \text{ pc} \\ &= \frac{1}{67.80 \times 10^3 \text{ ms}^{-1}} \times 10^6 \times 3.09 \times 10^{16} \text{ m} = 4.557 \times 10^{17} \text{ s} \\ &= \frac{4.557 \times 10^{17} \text{ s}}{365 \times 24 \times 60 \times 60 \text{ s yr}^{-1}} = 14.5 \times 10^9 \text{ yr} \end{aligned}$$

This assumes that the universe has been expanding at a constant rate. This is not the case, and so this is an overestimate. The actual age of the universe according to the data from the Planck satellite is 13.8 billion years.

D3.4 The cosmic microwave background radiation

In 1964, Arno Penzias and Robert Wilson, two radio astronomers working at Bell Laboratories, made a fundamental, if accidental, discovery. They were using an antenna they had just designed to study radio signals from our galaxy. But the antenna was picking up a signal that persisted no matter what part of the sky the antenna was pointing at. The spectrum of this signal (that is, the amount of energy as a function of the wavelength) turned out to be a black-body spectrum

Exam tip

The inverse of the Hubble constant gives an **upper bound** on the age of the universe – that is, the actual age is less. This is because the estimate is based on a constant rate of expansion equal to the present rate.

Exam tip

The characteristics of the cosmic microwave background are:

- a spectrum corresponding to black-body radiation at a temperature of 2.7 K.
- peak radiation in the microwave region.
- isotropic radiation with no apparent source.



corresponding to a temperature of 2.7 K. The **isotropy** of this radiation (the fact that it was the same in all directions) indicated that it was not coming from any particular spot in the sky; rather, it was radiation that was filling all space.

Penzias and Wilson did not know that this kind of radiation had been predicted on the basis of the Big Bang theory 30 years earlier by George Gamow and his co-workers, and more recently by Jim Peebles and Robert Dicke at Princeton. The Princeton group was in fact planning to start a search for this radiation when the news of the discovery arrived.

Penzias and Wilson, with help from the Princeton group, realised that the radiation detected was the remnant of the hot explosion at the beginning of time. It was the afterglow of the enormous temperatures that existed in the very early universe. As the universe has expanded, the temperature has fallen to its present value of 2.7 K.

Since the work of Penzias and Wilson, a number of satellite observatories – COBE (COsmic Background Explorer), WMAP (Wilkinson Microwave Anisotropy Probe) and the Planck satellite – have verified the black-body nature of this cosmic microwave background radiation to extraordinary precision and measured its present temperature to be 2.723 K.



The COBE collaboration was headed by John Mather, who was in charge of over a thousand scientists and engineers.

Worked example

D.17 Find the wavelength at which most CMB radiation is emitted.

From the Wien displacement law, $\lambda T = 2.9 \times 10^{-3} \text{ K m}$, it follows that most of the energy is emitted at a wavelength of $\lambda = 1.07 \text{ mm}$, which is in the microwave region.

D3.5 The accelerating universe and red-shift

It was expected that the rate of expansion of the universe should be slowing down. This was a reasonable expectation based on the fact that gravity should be pulling back on the distant galaxies, slowing them down. Cosmologists were very much interested in determining the value of the **deceleration parameter** of the universe, a dimensionless number called q_0 that would quantify the deceleration. A positive value of q_0 would indicate deceleration and a slowing down of the expansion rate. No one doubted that q_0 would be positive; the only issue was its actual value.

Two groups started work in this direction. The first group, under Saul Perlmutter, started the search in 1988 and the other, led by Brian Schmidt and Adam Riess, started in 1994.

We mentioned earlier that the distance–red-shift relation, $d = \frac{cz}{H_0}$, is only approximate, valid for $z < 0.2$. But the two groups were looking for very distant supernovae and therefore high z values. A more accurate relation in this case is

$$d = \frac{cz}{H_0} \left(1 + \frac{1}{2}(1 - q_0)z \right)$$

Exam tip

You will not be examined on the formula

$$d = \frac{cz}{H_0} \left(1 + \frac{1}{2}(1 - q_0)z \right)$$

but knowing of its existence is crucial in understanding how the conclusion of an **accelerating universe** was reached.

We see that the deceleration parameter q_0 appears in this formula. So, *in theory*, the task looked easy: find distant objects, measure their distance d and red-shift z and use the data to determine q_0 . The two groups chose to look at distant **Type Ia supernovae**. The nature of these will be discussed in more detail in Section **D4.4**.

Type Ia supernovae are very rare: only a few would be expected to occur in a galaxy every thousand years! The great discovery about Type Ia supernovae is that they all have the same peak luminosity and so may be used as standard candles. Figure **D.27** shows how the logarithm of the luminosity (in W) of a Type Ia supernova varies with time.

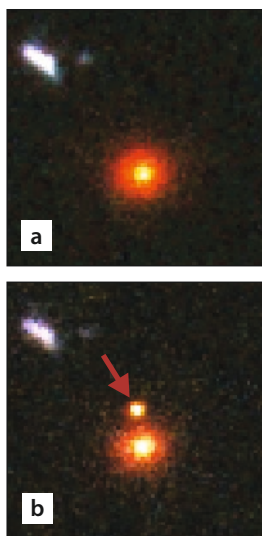


Figure D.28 Observation of a Type Ia supernova, **a** before and **b** after outburst.

Exam tip

A ‘standard candle’ refers to a star of known luminosity. Thus measuring its apparent brightness gives its distance.

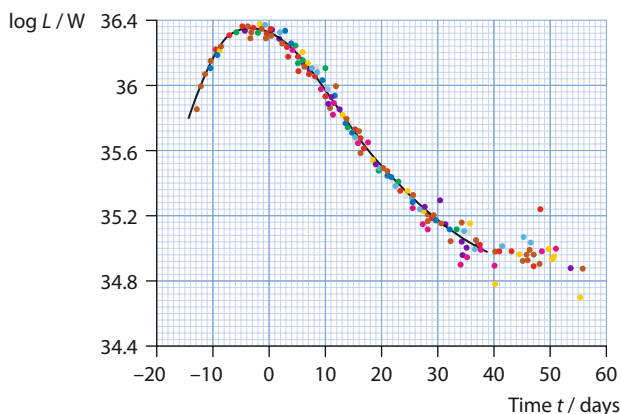


Figure D.27 Different Type Ia supernovae all have the same peak luminosity. (Adapted from graph ‘Low redshift type 1a template lightcurve’, Supernova Cosmology Project/Adam Riess)

The peak luminosity is a staggering 2×10^{36} W and falls off with time over a couple of months. If one is lucky enough to observe a Type Ia supernova from before the peak luminosity is reached, the peak apparent brightness b may be measured, and since the peak luminosity L is known we may find the distance to the supernova using $b = \frac{L}{4\pi d^2}$.

The task which looked easy in theory was formidable in practice, but a total of 45 supernovae were studied by the first group and 16 by the second (Figure **D.28**). Both groups were surprised to find that the deceleration parameter came out negative. This meant that the rate at which distant objects are moving away from us is **increasing**: the universe is **accelerating**. Perlmutter, Schmidt and Riess shared the 2011 Nobel Prize in Physics for this extraordinary discovery.

Worked example

D.18 Show that at a temperature $T = 10^{10}$ K there is enough thermal energy to create electron–positron pairs.

The thermal energy corresponding to $T = 10^{10}$ K is $E_k \approx \frac{3}{2}kT = 1.5 \times 1.38 \times 10^{-13} \text{ J} = 1.3 \text{ MeV}$. The **rest energy** of an electron is $m_e c^2 \approx 0.5 \text{ MeV}$, so the thermal energy $E_k \approx 1.3 \text{ MeV}$ is enough to produce a pair.



Nature of science

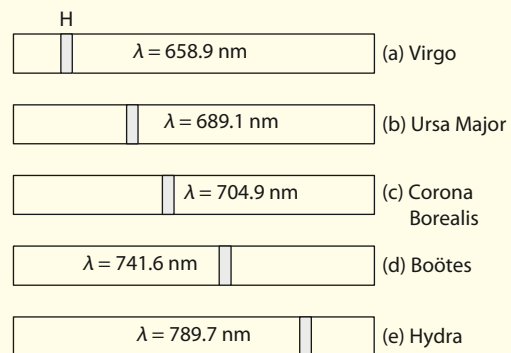
Occam's razor

Any theory of the origin of the universe must fit with the data available. The principle of **Occam's razor** says that the simplest explanation is likely to be the best. In the debate between conflicting cosmological theories, an expanding universe and the existence of a background radiation were predicted by the Big Bang theory. The red-shift in the light from distant galaxies was evidence for expansion, but it was the discovery of the cosmic microwave background radiation that led to the acceptance of the Big Bang theory. Although this theory explains many aspects of the universe as we see it now, it cannot explain what happened at the very instant the universe was created.

? Test yourself

- 43 **a** State and explain **Hubble's law**.
- b** Explain how this law is evidence for an expanding universe.
- 44 Some galaxies actually show a blue-shift, indicating that they are moving towards us. Discuss whether this violates Hubble's law.
- 45 Galaxies are affected by the gravitational pull of neighbouring galaxies and this gives rise to what are called **peculiar** velocities. Typically these are about 500 km s^{-1} . Estimate how far away a galaxy should be so that its velocity of recession due to the expanding universe equals its peculiar velocity.
- 46 A student explains the expansion of the universe as follows: 'Distant galaxies are moving at high speeds into the vast expanse of empty space.' Suggest what is wrong with this statement.
- 47 It is said that the Big Bang started everywhere in space. Suggest what this means.
- 48 In the context of the Big Bang theory, explain why the question 'what existed before the Big Bang?' is meaningless.
- 49 Suppose that at some time in the future a detailed study of the Andromeda galaxy and all the nearby galaxies in our Local Group will be possible. Discuss whether this would help in determining Hubble's constant more accurately.
- 50 The diagram shows two lines due to calcium absorption in the spectra of five galaxies, ranging from the nearby Virgo to the very distant Hydra. Each diagram gives the wavelength of the H (hydrogen) line. The wavelength of the H line in the lab is 656.3 nm .

Using Hubble's law, find the distance to each galaxy. (Use $H = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.)



- 51 Take Hubble's constant H at the present time to be $68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
 - a** Estimate at what distance from the Earth the speed of a receding galaxy is equal to the speed of light.
 - b** Suggest what happens to galaxies that are beyond this distance.
 - c** The theory of special relativity states that nothing can exceed the speed of light. Suggest whether the galaxies in **b** violate relativity.
- 52 Discuss **three** pieces of evidence that support the Big Bang model of the universe.
- 53 A particular spectral line, when measured on Earth, corresponds to a wavelength of $4.5 \times 10^{-7} \text{ m}$. When received from a distant galaxy, the wavelength of the same line is measured to be $5.3 \times 10^{-7} \text{ m}$.
 - a** Calculate the red-shift for this galaxy.
 - b** Estimate the speed of this galaxy relative to the Earth.
 - c** Estimate the distance of the galaxy from the Earth. (Take $H = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.)

- 54 Explain why the inverse of the Hubble constant, $\frac{1}{H}$, is taken to be an estimate of the 'age of the universe'. Estimate how old the universe would be if $H = 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (close to Hubble's original value).
- 55 Explain why the Hubble constant sets an **upper bound** on the age of the universe.
- 56 Explain why Hubble's law does not imply that the Earth is at the centre of the universe.
- 57 The temperature of the cosmic microwave background radiation as measured from the Earth is about 2.7 K.
- What is the significance of this radiation?
 - What would be the temperature of the CMB radiation as measured by an observer in the Andromeda galaxy, 2.5 million light years away?
- 58 **a** Draw a sketch graph to show the variation of the CMB radiation intensity with wavelength.
- Calculate the peak wavelength corresponding to a CMB radiation temperature of 2.72 K.
- 59 Predict what will happen to the temperature of the CMB radiation if:
- the universe keeps expanding forever
 - the universe starts to collapse.
- 60 **a** State what is meant by **red-shift**.
- Describe the mechanism by which the observed red-shift in light from distant galaxies is formed.
 - Show that the distance d of a galaxy with a red-shift of z is given by $d = \frac{cz}{H_0}$.
 - Calculate the distance of a galaxy whose red-shift is 0.18, using $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
 - Estimate the size of the universe now, relative to its size when the light in **d** was emitted.
- 61 The wavelength of a particular spectral line measured in the laboratory is 486 nm. The same line observed in the spectrum of a distant galaxy is shifted by 15 nm.
- Estimate the distance of the galaxy, using $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
 - Estimate the size of the universe now, relative to its size when the light in **a** was emitted.
- 62 A photon is emitted at a time when the size of the universe was 85% of its present size. Estimate the distance from the Earth of the point from which the photon was emitted, using $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
- 63 State the property of Type Ia supernovae that is significant in distance measurements in cosmology.
- 64 **a** State what is meant by the **accelerating universe**.
- Suggest why the universe was expected to decelerate rather than accelerate.
 - Outline how Type Ia supernovae were used to discover the acceleration of the universe.
 - Explain why it is important to observe such a supernova starting before it reaches its peak luminosity.
- 65 It is said that distant supernovae appear dimmer than they would in a decelerating universe. Explain this statement.
- 66 It was stated in the text that Type Ia supernovae are very rare (a few in a galaxy every thousand years). Suggest how two research groups were able to observe over 50 such supernovae in the space of just a few years.

Learning objectives

- Understand and apply the Jeans criterion.
- Describe nuclear fusion in stars.
- Describe nucleosynthesis off the main sequence.
- Distinguish and describe Type Ia and Type II supernovae.

D4 Stellar processes (HL)

This section deals with the birth, evolution and death of stars, and with their role in **nucleosynthesis**, the production of elements through fusion and neutron absorption. The section closes with a discussion of supernovae and the role of Type Ia supernovae as standard candles.

D4.1 The Jeans criterion

Interstellar space (the space between stars) consists of gas and dust at a density of about $10^{-21} \text{ kg m}^{-3}$. This amounts to about one atom of hydrogen in every cubic centimetre of space. The gas is mainly hydrogen (about 74% by mass) and helium (25%), with other elements making up



the remaining 1%. Whenever the gravitational energy of a given mass of gas exceeds the average kinetic energy of the random thermal motion of its molecules, the gas becomes unstable and tends to collapse:

$$\frac{GM^2}{R} \geq \frac{3}{2} NkT$$

where k is Boltzmann's constant, T is temperature and N is the number of particles. R is the radius of the gas cloud and M its mass. This is known as the **Jeans criterion**.

Stars formed (and continue to be formed) when rather cool gas clouds in the interstellar medium ($T \approx 10\text{--}100\text{K}$) of sufficiently large mass (large enough to satisfy the Jeans criterion) collapsed under their own gravitation. In the process of contraction, the gas heated up. Typically, the collapsing gas would break up into smaller clouds, resulting in the creation of more than one star. When the temperature rises sufficiently for visible light to be emitted, a star so formed is called a **protostar** (see Figure D.29).

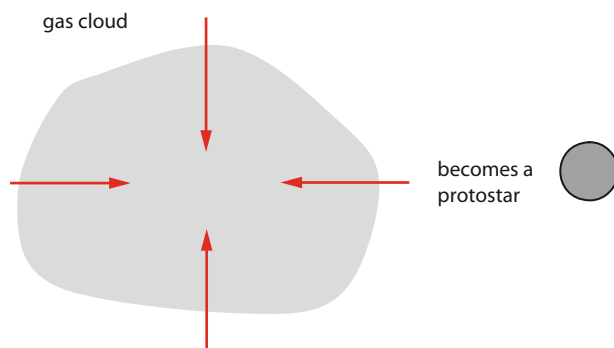


Figure D.29 The formation of a protostar out of a collapsing cloud of gas.

Worked examples

D.19 Show that the Jeans criterion can be rewritten as $M^2 = \frac{3}{4\pi\rho} \left(\frac{3kT}{2mG} \right)^3$, where ρ is the density of the gas and m is the mass of a particle of the gas. (The right hand side is known as the square of the Jeans mass.)

Cube each side of the Jeans criterion equation to find

$$\left(\frac{GM^2}{R} \right)^3 = \left(\frac{3kMT}{2m} \right)^3 \Rightarrow M^2 = \left(\frac{3}{4\pi\rho} \right) \left(\frac{3kT}{2mG} \right)^3$$

using the definition of density and $M = Nm$, where m is the mass of one molecule.

D.20 Take the density of interstellar gas in a cloud to be about 100 atoms of hydrogen per cm^3 . Estimate the smallest mass this cloud can have for it to become unstable and begin to collapse when $T = 100\text{K}$.

The density is

$$\frac{100 \times 1.67 \times 10^{-27} \text{kg}}{10^{-6} \text{m}^3} = 1.67 \times 10^{-19} \text{kgm}^{-3}$$

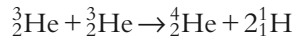
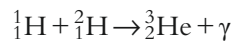
With $T = 100\text{K}$ in the Jeans criterion we find (see Worked example D.19)

$$M \approx 3.0 \times 10^{33} \text{kg} = 1.5 \times 10^3 M_{\odot}$$

Such a large gas cloud might well break up, forming more than one star.

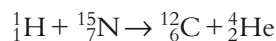
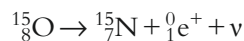
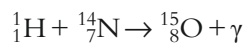
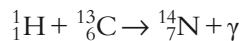
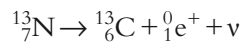
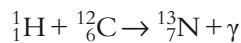
D4.2 Nuclear fusion

We saw in Section D1.2 that the main series of nuclear fusion reactions taking place in the cores of main-sequence stars is the proton–proton cycle:



In this cycle, the net effect is to turn four hydrogen nuclei into one helium nucleus.

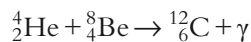
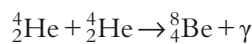
For stars more massive than our Sun, there is a second way to fuse hydrogen into helium. This is the so-called **CNO cycle**, described by the following series of fusion reactions:



Notice that the net effect is to turn four hydrogen nuclei into one helium nucleus, just like the proton–proton cycle; the heavier elements produced in intermediate stages are all used up. The carbon nucleus has a charge of +6, so the barrier that must be overcome for carbon to fuse is much higher. This requires higher temperatures.

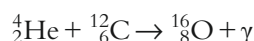
The proton–proton cycle and the CNO cycle are both main-sequence star processes. What happens beyond the main sequence?

The first element to be produced as a star leaves the main sequence and enters the red giant stage is carbon, through the **triple alpha process**:

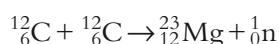
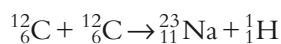
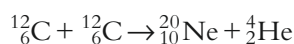


This happens to stars with masses up to eight solar masses. The star will shed most of its mass in a planetary nebula and end up as a white dwarf with a core of carbon.

For stars even more massive than this, helium fuses with carbon to produce oxygen:



In even more massive stars, neon, sodium and magnesium are produced:



Exam tip

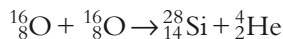
The CNO cycle applies to more massive main-sequence stars and does not produce elements heavier than helium.

Exam tip

The triple alpha process is how we end up with white dwarfs with a carbon core.



Silicon is then produced by the fusion of oxygen:



The process continues until iron is formed. This creates an onion-like layered structure in the star, with progressively heavier elements as we move in towards the centre. Fusion cannot produce elements heavier than iron, since the binding energy per nucleon peaks near iron and further fusion is not energetically possible. Thus, a massive star ends its cycle of nuclear reactions with iron at its core, surrounded by progressively lighter elements, as shown in Figure D.30.

Exam tip

You should be able to explain why fusion ends with the production of iron.

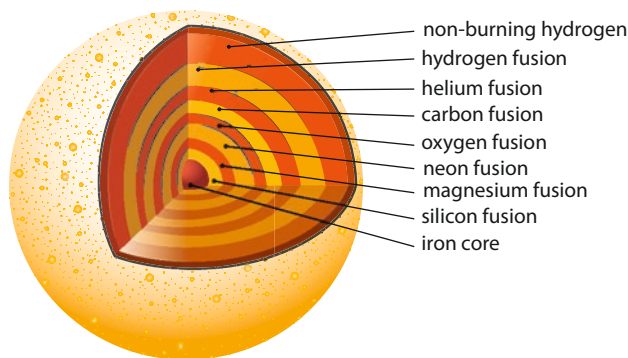


Figure D.30 The central core of a fully evolved massive star consists of iron with layers of lighter elements surrounding it.

Worked example

D.21 Our Sun emits energy at a rate (luminosity) of about 3.9×10^{26} W. Estimate the mass of hydrogen that undergoes fusion in one year. Assuming that the energy loss is maintained at this rate, find the time required for the Sun to convert 12% of its hydrogen into helium. (Mass of Sun = 1.99×10^{30} kg.)

Assuming the proton–proton cycle as the reaction releasing energy (by hydrogen fusion), the energy released per reaction is about 3.98×10^{-12} J. Since the luminosity of the Sun is 3.9×10^{26} W, it follows that the number of fusion reactions required per second is

$$\frac{3.9 \times 10^{26}}{3.98 \times 10^{-12}} = 9.8 \times 10^{37}$$

For every such reaction, four hydrogen nuclei fuse to helium and thus the mass of consumed hydrogen is $9.8 \times 10^{37} \times 4 \times 1.67 \times 10^{-27} \text{ kg s}^{-1} = 6.5 \times 10^{11} \text{ kg s}^{-1} = 2 \times 10^{19} \text{ kg}$ per year. At the time of its creation, the Sun consisted of 75% hydrogen, corresponding to a mass of $0.75 \times 1.99 \times 10^{30} \text{ kg} = 1.5 \times 10^{30} \text{ kg}$. The limit of 12% results in a hydrogen mass to be fused of $1.8 \times 10^{29} \text{ kg}$. The time for this mass to fuse is thus

$$\begin{aligned} \frac{1.8 \times 10^{29}}{6.5 \times 10^{11} \text{ s}} &= 2.8 \times 10^{17} \text{ s} \\ &= 8.9 \times 10^9 \text{ yr} \end{aligned}$$

Since the Sun has existed for about 5 billion years, it still has about 4 billion years left in its life as a main-sequence star.

One application of the mass–luminosity relation is to estimate the lifetime of a star on the main sequence. Since the luminosity is the power radiated by the star, we may write that

$$\frac{E}{T} \propto M^{3.5}$$

where E is the total energy radiated by the star and T is the time in which this happens. For the purposes of an estimate, we may assume that the total energy that the star can radiate comes from converting *all* its mass into energy according to Einstein’s formula, $E = Mc^2$. Thus

$$\frac{E}{T} \propto M^{3.5} \Rightarrow \frac{Mc^2}{T} \propto M^{3.5} \Rightarrow T \propto M^{-2.5}$$

This means that the lifetimes of two stars are approximately related by

$$\frac{T_1}{T_2} = \left(\frac{M_2}{M_1}\right)^{2.5}$$

Worked example

D.22 Our Sun will spend about 10^{10} yr on the main sequence. Estimate the time spent on the main sequence by a star whose mass is 10 times the mass of the Sun.

We know that $\frac{T_1}{T_2} = \left(\frac{M_2}{M_1}\right)^{2.5}$. Hence, $\frac{T}{T_\odot} = \left(\frac{M_\odot}{10M_\odot}\right)^{2.5} \Rightarrow T = \frac{10^{10}}{10^{2.5}} \approx 3 \times 10^7$ yr.

D4.3 Nucleosynthesis of the heavy elements

All of the hydrogen and most of the helium in the universe were produced at the very earliest moments in the life of the universe. Everything else was made in stars in the course of stellar evolution. In Section **D4.2** we learned that nuclear fusion reactions in stellar cores produce the elements up to iron. So how are the rest of the elements in the periodic table produced?

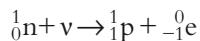
The answer lies in neutron absorption by nuclei – **neutron capture**. When a nucleus absorbs a neutron it becomes an isotope of the original nucleus. This isotope is usually unstable and will decay. The issue is whether there is enough time for this decay to occur before the isotope absorbs yet another neutron. In what is referred to as an **s-process** (s for ‘slow’), the isotope does have time to decay because the number of neutrons present is small. This happens in stars where relatively small numbers of neutrons are produced in the fusion reactions discussed in Section **D4.2**. The isotope will undergo a series of decays, including beta decay, in which the atomic number is increased by one, thus producing a new element. This process accounts for the production of about half of the nuclei above iron, and ends with the production of bismuth 209.

By contrast, in the presence of very large numbers of neutrons, nuclei that absorb neutrons do not have time to decay. In an **r-process** (r for ‘rapid’), they keep absorbing neutrons one by one, forming very heavy,



neutron-rich isotopes. This cannot happen inside a star but it does happen during a supernova explosion. These neutron-rich isotopes are then hurled into space by the supernova, where they can undergo beta decay, producing nuclei of higher atomic number.

Beta decay is not the only way to turn a neutron into a proton and hence increase the atomic number. In supernova explosions, massive numbers of neutrinos are produced. A neutron may absorb a neutrino and turn into a proton according to the reaction



D4.4 Type Ia and Type II supernovae

The supernovae we learned about in Section D2.8 referred to the explosion of red supergiant stars. These are called **Type II supernovae**. Another type of supernovae, called **Type Ia supernovae**, involve a different mechanism.

Consider a **binary star** system in which one of the stars is a white dwarf. This star may attract material from its companion star. Mass falling into the white dwarf may increase the white dwarf's mass beyond the Chandrasekhar limit, so nuclear fusion reactions may start again in its core. The resulting sudden release of energy appears as a sudden increase in the luminosity of the white dwarf – that is, a supernova. Unlike Type II supernovae, Type Ia supernovae show no hydrogen lines.

Figure D.31 shows the remnants of two supernovae.

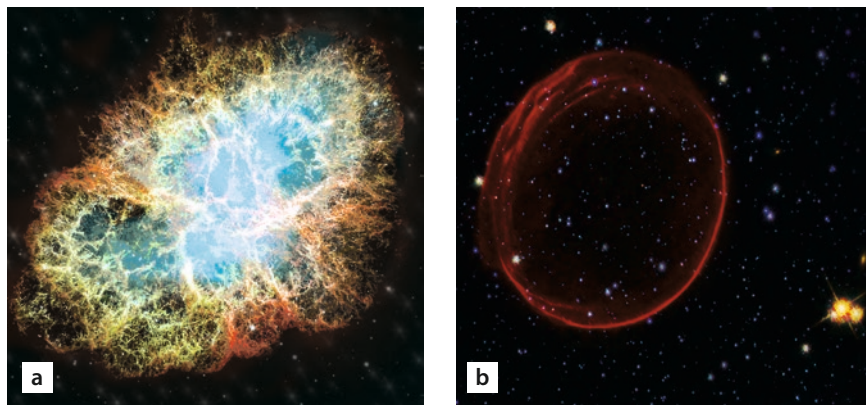


Figure D.31 **a** The Crab nebula, the remnant of a Type II supernova observed by the Chinese in 1054 and possibly by Native Americans. It was visible for weeks, even during daytime. **b** This majestically serene bubble of gas in space is the remnant of the Type Ia supernova SNR 0509, which exploded about 400 years ago.

As noted in Section D3.5, an important property of Type Ia supernovae is that they all have the same peak luminosity and so may be used as 'standard candles'. By measuring their apparent brightness at the peak, we may calculate their distance from $d = \sqrt{\frac{L}{4\pi b}}$ (see Section D1.5).

Apart from the mechanism producing them, the two types of supernovae also differ in that Type Ia supernovae do not have hydrogen absorption lines in their spectra, whereas Type II do. They also differ in the way the luminosity falls off with time (Figure D.32).

Exam tip

Differences in the two types of supernovae:

Type Ia

- do not have hydrogen lines in their spectra
- are produced when mass from a companion star accretes onto a white dwarf, forcing it to exceed the Chandrasekhar limit
- have a luminosity which falls off sharply after the explosion.

Type II

- have hydrogen lines in their spectra
- are produced when a massive red supergiant star explodes
- have a luminosity which falls off gently after the explosion.

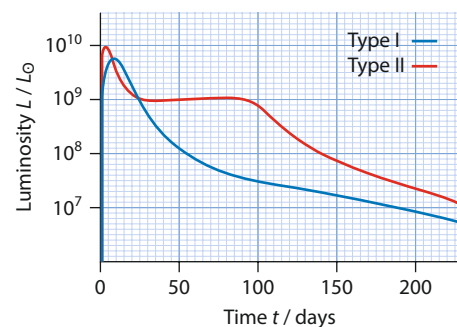


Figure D.32 Type Ia and Type II supernovae show different graphs of luminosity versus time.

Nature of science

Observation and deduction

Stellar spectra show us the elements present in stars' outer layers, but how do we explain how these elements came to be formed?

Astrophysicists have shown how nuclear fusion reactions in the stars produce energy and also build up the elements. Modelling and computer simulations have resulted in a picture that agrees very well with observation. The time scales of stellar processes are much too long for us to directly observe the phenomena of stellar evolution, but the presence of stars of different ages and at different stages of evolution allows a comparison of observations and theoretical predictions.

? Test yourself

- 67 Describe the **Jeans criterion**.
- 68 Describe how a gas and dust cloud becomes a protostar.
- 69 Explain whether star formation is more likely to take place in cold or hot regions of interstellar space.
- 70 Explain why a star at the top left of the main sequence will spend much less time on the main sequence than a star at the lower right.
- 71 Show that a star that is twice as massive as the Sun has a lifetime that is 5.7 times shorter than that of the Sun.
- 72 Using the known luminosity of the Sun and assuming that it stays constant during the Sun's lifetime, which is estimated to be 10^{10} yr, calculate the mass this energy corresponds to according to Einstein's mass–energy formula.
- 73 Suggest why the depletion of hydrogen in a star is such a significant event.
- 74 Evolved stars that have left the main sequence have an onion-like layered structure. Outline how this structure is created.
- 75 Describe the nuclear reactions taking place in a star of one solar mass:
 - a while the star is on the main sequence
 - b after it has left the main sequence.
- 76 Describe the nuclear reactions taking place in a star of 20 solar masses:
 - a while on the main sequence
 - b after it has left the main sequence.
- 77 Explain why no elements heavier than iron are produced in stellar cores.
- 78 State the element that is the end product of:
 - a the proton–proton cycle
 - b the CNO cycle
 - c the triple alpha process.
- 79 Distinguish between an **s-process** and an **r-process**.
- 80 Suggest why the production of heavier elements inside stars requires higher temperatures.
- 81 Describe how a Type Ia supernova is formed.
- 82 Describe how a Type II supernova is formed.
- 83 State **three** differences between Type Ia and Type II supernovae.
- 84 Suggest why hydrogen lines are expected in the spectra of Type II supernovae.
- 85 Compare and contrast the proton–proton and CNO cycles.



D5 Further cosmology (HL)

This section deals with some open questions in cosmology, questions that are the subject of intensive current research. These include the evidence for and the nature of dark matter and dark energy, fluctuations in the CMB, and the **rotation curves** of galaxies

D5.1 The cosmological principle

The universe appears to be full of structure. There are planets and moons in our solar system, there are stars in our galaxy, our galaxy is part of a **cluster of galaxies** and our cluster is part of an even bigger **supercluster** of galaxies.

If we look at the universe on a very large scale, however, we no longer see any structure. If we imagine cutting up the universe into cubes some 300 Mpc on a side, the interior of any one of these cubes would look much the same as the interior of any other, anywhere else in the universe. This is an expression of the so-called **homogeneity principle** in cosmology: on a large enough scale, the universe looks uniform.

Similarly, if we look in different directions, we see essentially the same thing. If we look far enough in any direction, we will count the same number of galaxies. No one direction is special in comparison with another. This leads to a second principle of cosmology, the **isotropy principle**. A related observation is the high degree of isotropy of the CMB.

These two principles, homogeneity and isotropy, make up what is called the **cosmological principle**, which has had a profound role in the development of models of cosmology.

The cosmological principle implies that the universe has no edge (for if it did, the part of the universe near the edge would look different from a part far from the edge, violating the homogeneity principle). Similarly, it implies that the universe has no centre (for if it did, an observation from the centre would show a different picture from an observation from any other point, violating the principle of isotropy).

D5.2 Fluctuations in the CMB

We have noted several times that the CMB is uniform and isotropic. However, it is not perfectly so. There are small variations ΔT in temperature, of the order of $\frac{\Delta T}{T} \approx 10^{-5}$, where $T = 2.723 \text{ K}$ is the average temperature. These variations in temperature are related to variations in the density of the universe. In turn, variations in density are the key to how structures formed in the universe. With perfectly uniform temperature and density in the universe, stars and galaxies would not form.

In addition to helping us understand structures, CMB anisotropies are related to the geometry of the universe. There would be different degrees of anisotropy depending on whether the universe has positive, zero or negative curvature (see the section on the dependence of the **scale factor** on time later in this section). A number of investigations of CMB anisotropies have been carried out, using COBE, WMAP, the Planck satellite observatory and the Boomerang (Balloon Observations

Learning objectives

- Describe the cosmological principle.
- Understand the fluctuations in the CMB.
- Understand the cosmological origin of red-shift.
- Derive the critical density and understand its significance.
- Describe dark matter.
- Derive rotation curves and understand how they provide evidence for dark matter.
- Understand dark energy.
- Sketch the variation of the scale factor with time for various models.

The anisotropies in the CMB are crucial in understanding the formation of structures.

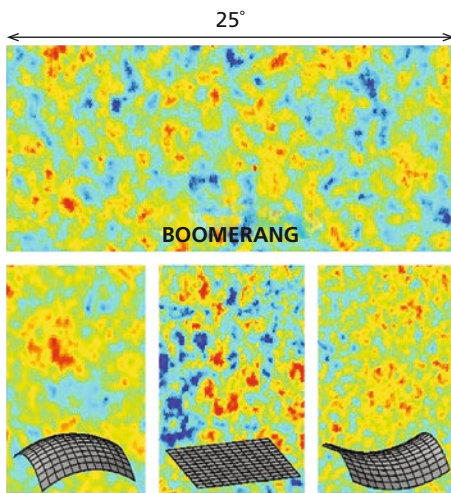


Figure D.33 Fluctuations in temperature are shown as differences in colour in this image from the Boomerang collaboration. Theoretical models using space of different curvatures are also shown. There is a clear match with the flat-universe case. © The Boomerang Collaboration.

Of Millimetric Extragalactic Radiation) collaboration. Figure **D.33** shows fluctuations in the CMB temperature obtained by the Boomerang collaboration, and three theoretical predictions of what that anisotropy should look like in models with positive, negative and zero **curvature** of space. Different colours correspond to different temperatures. Even judged by eye, the data appear to be consistent with the flat case.

Figure **D.34** is a spectacular map from the Planck satellite observatory, showing **CMB fluctuations** in temperature as small as a few millionths of a degree. This is a map of the radiation filling the universe when it was only about 380 000 years old.

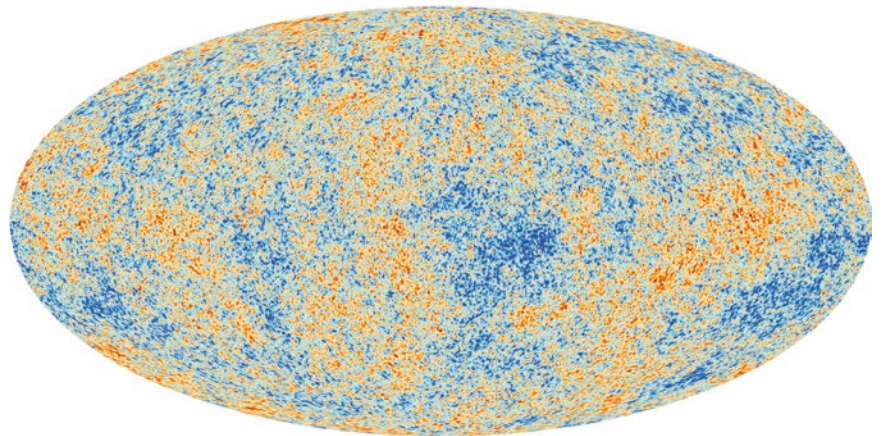


Figure D.34 Fluctuations in temperature of the CMB according to ESA's Planck satellite observatory. (©ESA and the Planck Collaboration, reproduced with permission)

Studies of CMB anisotropy also give crucial information on cosmological parameters such as the density of matter and energy in the universe.

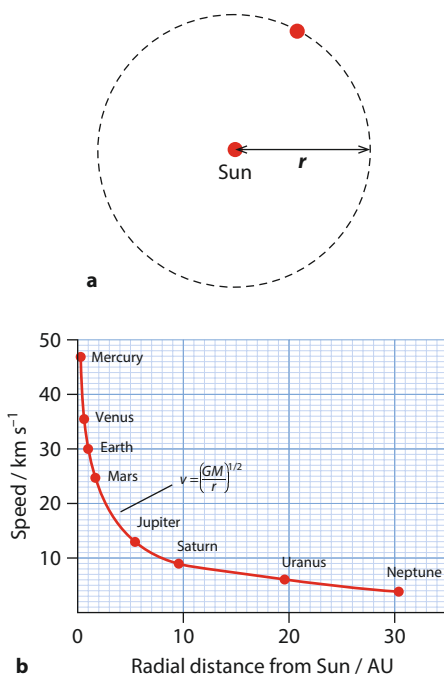


Figure D.35 **a** A particle orbiting a central mass. **b** The rotation curve of the particle in **a** shows a characteristic drop. (From M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004. © The Open University, used with permission)

D5.3 Rotation curves and the mass of galaxies

Consider a planet as it revolves about the Sun (Figure **D.35a**). In Topics **6** and **10** we determined, using $\frac{GMm}{r^2} = \frac{mv^2}{r}$, that the speed of a planet a distance r from the Sun is $v = \sqrt{\frac{GM}{r}}$. This means that $v \propto \frac{1}{\sqrt{r}}$. Plotting rotational speed against distance gives what is called a **rotation curve**, as shown in Figure **D.35b**.

Now consider a spherical mass cloud of uniform density (Figure **D.36a**). What is the speed of a particle rotating about the centre at a distance r ? We can still use $v = \sqrt{\frac{GM}{r}}$, but now M stands for the mass in the spherical body up to a distance r from the centre. Since the density is constant we have that

$$\rho = \frac{M}{V} = \frac{M}{\frac{4\pi r^3}{3}} = \frac{3M}{4\pi r^3}$$

and hence

$$M = \frac{4\pi r^3 \rho}{3}$$



Thus

$$v = \sqrt{\frac{G4\pi r^3 \rho}{3r}}$$

and so $v \propto r$. The rotation curve is a straight line through the origin. This is valid for r up to R , the radius of the spherical mass cloud. Beyond R the curve is like that of Figure D.35.

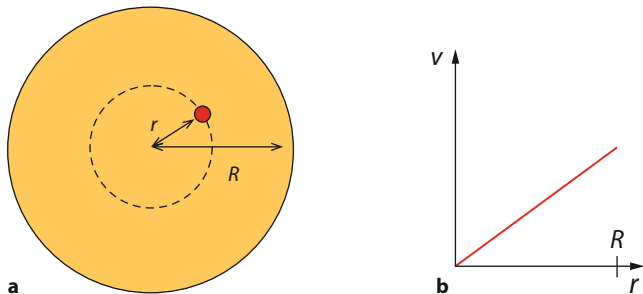


Figure D.36 **a** A particle orbiting around the centre of a uniform spherical cloud. **b** The rotation curve of the particle in **a**.

Worked example

D.23 Consider a system in which the mass varies with distance from the axis according to $M = kr$, where k is a constant. Derive the rotation curve for such a system.

We start with $v = \sqrt{\frac{GM}{r}}$. We get $v = \sqrt{\frac{Gkr}{r}} = \sqrt{Gk} = \text{constant}$. The rotation curve would thus be a horizontal straight line.

Figure D.37 shows the rotation curve of the Milky Way galaxy.

This rotation curve is not one that belongs to a large central mass, as in Figure D.35. Its main feature is the flatness of the curve at large distances from the centre. Notice that the flatness starts at distances of about 13 kpc. The central galactic disc has a radius of about 15 kpc. This means that there is substantial mass outside the central galactic disc. Furthermore, according to Worked example D.23, a flat curve corresponds to increasing mass with distance from the centre.

The flat part of the galaxy's rotation curve indicates substantial mass far from the centre.

If the mass were contained within a given distance, then past that distance the rotation curve should have dropped, as it does in Figure D.35. The problem is that no such drop is seen – but at the same time no such mass is visible. Arguments like this have led to the conclusion that there must be considerable mass at large distances past the galactic disc. This is **dark matter**: matter that is too cold to radiate and so cannot be seen. It is estimated that, in our galaxy, dark matter forms a spherical halo around the galaxy and has a mass that is about 10 times larger than the mass of all the stars in the galaxy!

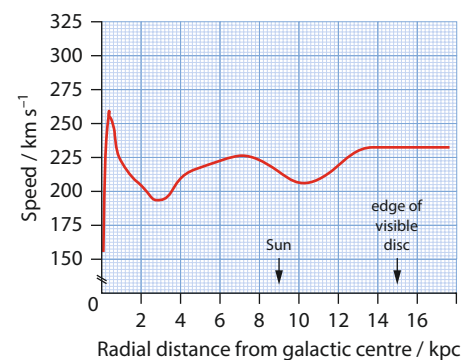


Figure D.37 The rotation curve of our galaxy shows a flat region, indicating the presence of matter far from the galactic disc. (Adapted from Combes, F. (1991) Distribution of CO in the Milky Way, Annual Review of Astronomy and Astrophysics, 29, pp.195–237)

D5.4 Dark matter

It is estimated that 85% of the matter in the universe is dark matter. It cannot be seen; we know of its existence mainly from its gravitational effects on nearby bodies.

What could dark matter be? It could be ordinary, cold matter that does not radiate – like, for example, brown dwarfs, black dwarfs or small planets. Collectively these are called MACHOs (MASSIVE Compact Halo Objects). This is matter consisting mainly of protons and neutrons, so it is also called **baryonic matter**. The problem is that there are limits to how much baryonic matter there can be. The limit is at most 15%, so dark matter must also contain other, more exotic forms.

The class of non-baryonic objects which are candidates for dark matter are called WIMPs (for Weakly Interacting Massive Particles). Neutrinos fall into this class since they are known to have a small mass, although their tiny mass is not enough to account for all non-baryonic dark matter. Unconfirmed theories of elementary particle physics based mainly on the idea of **supersymmetry** (a proposed symmetry between particles with integral spin and particles with half-integral spin) predict the existence of various particles that would be WIMP candidates – but no such particles have been discovered.

So the answer to the question ‘what is dark matter?’ is mainly unknown at the moment.

D5.5 The cosmological origin of red-shift

In Section D3.2 we derived the formula

$$\frac{\lambda}{\lambda_0} = \frac{R}{R_0}$$

where R_0 is the value of the scale factor at the time of emission of a photon of wavelength λ_0 , R is the value of the scale factor at the present time (when the photon is received) and λ is the wavelength of the photon as measured at the present time. This gives a cosmological interpretation of the red-shift, rather than one based on the Doppler effect: the space in between us and the galaxies is stretching, so wavelengths stretch as well.

A direct consequence of this is on the temperature of the CMB radiation that fills the universe. The wavelength λ_0 corresponds to a CMB temperature of T_0 . By the Wien displacement law,

$$\lambda_0 T_0 = \lambda T = \text{constant}$$

Therefore

$$\frac{\lambda}{\lambda_0} = \frac{T_0}{T}$$

This implies that

$$\frac{T_0}{T} = \frac{R}{R_0} \text{ or } T \propto \frac{1}{R}$$

This shows that, as the universe expands (that is, as R gets bigger), the temperature drops. This is why the universe is cooling down, and why the present temperature of the CMB is so low (2.7 K).



Worked example

D.24 The photons of CMB radiation observed today are thought to have been emitted at a time when the temperature of the universe was about 3.0×10^3 K. Estimate the size of the universe then compared with its size now.

From $T \propto \frac{1}{R}$ we find $\frac{T_0}{T} = \frac{R}{R_0}$, so $\frac{R}{R_0} = \frac{3.0 \times 10^3}{2.7} \approx 1100$, so the universe then was about 1100 times smaller.

One of the great problems in cosmology is to determine how the scale factor depends on time. We will look at this problem in the next two sections.

D5.6 Critical density

We begin the discussion in this section by solving a problem in Newtonian gravitation. Consider a spherical cloud of dust of radius r and mass M , and a mass m at the surface of this cloud which is moving away from the centre with a velocity v satisfying Hubble's law, $v = H_0 r$; see Figure D.38.

The total energy of the mass m is

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

If ρ is the density of the cloud, then $M = \rho \frac{4}{3}\pi r^3$. Using this together with $v = H_0 r$, we find

$$E = \frac{1}{2}mr^2 \left(H_0^2 - \frac{8\pi\rho G}{3} \right)$$

The mass m will continue to move away if its total energy is positive. If the total energy is zero, the expansion will halt at infinity; if it is negative, contraction will follow the expansion. The sign of the term $\left(H_0^2 - \frac{8\pi\rho G}{3} \right)$, that is, the value of ρ relative to the quantity

$$\rho_c = \frac{3H_0^2}{8\pi G} \approx 10^{-26} \text{ kg m}^{-3}$$

determines the long-term behaviour of the cloud.

This quantity is known as the **critical density**. We have only derived it in a simple setting based on Newton's gravitation. What does it have to do with the universe? We discuss this in the next section.

D5.7 The variation of the scale factor with time

In 1915 Einstein published his general theory of relativity, replacing Newton's theory of gravitation with a revolutionary new theory in which the rules of geometry are dictated by the distribution of mass and energy in the universe. Applying Einstein's theory to the universe as a whole results in equations for the scale factor R , and solving these gives the dependence of R on time. Einstein himself believed in a static universe – that is, a universe with $R = \text{constant}$. His equations, however, did not give a constant R . So he modified them, adding his famous **cosmological constant** term, Λ , to make R constant (see Figure D.39).

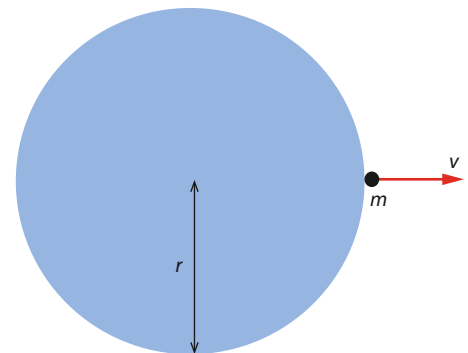


Figure D.38 Estimating critical density.

Exam tip

You must be able to derive the formula for the critical density.

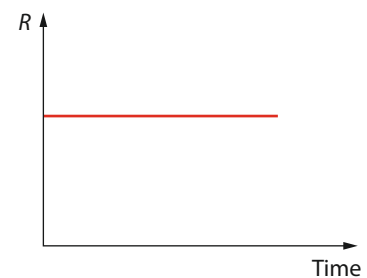


Figure D.39 A model with a constant scale factor. Einstein introduced the cosmological constant Λ in order to make the universe static.

In this model there is no Big Bang, and the universe always has the same size. This was before Hubble discovered the expanding universe. Einstein missed the great chance of theoretically predicting an expanding universe before Hubble; he later called adding the cosmological constant ‘the greatest blunder of [his] life’. This constant may be thought of as related to a ‘vacuum energy’, energy that is present in all space. The idea fell into obscurity for many decades but it did not go away: it was to make a comeback with a vengeance much later! It is now referred to as **dark energy**.

The first serious attempt to determine how R depends on time was made by the Russian mathematician Alexander Friedmann (1888–1925). Friedmann applied the Einstein equations and realised that there were a number of possibilities: the solutions depend on how much matter and energy the universe contains.

We define the **density parameters** Ω_m and Ω_Λ , for matter and dark energy respectively, as

$$\Omega_m = \frac{\rho_m}{\rho_c} \text{ and } \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c}$$

where ρ_m is the actual density of matter in the universe, ρ_c is the critical density derived in Section D5.6 and ρ_Λ is the density of dark energy. The Friedmann equations give various solutions depending on the values of Ω_m and Ω_Λ . Deciding which solution to pick depends crucially on these values, which is why cosmologists have expended enormous amounts of energy and time trying to accurately measure them.

Figure D.40 is a schematic representation of the possibilities; the subscript 0 indicates the values of these parameters at the present time. There are four regions in the diagram. The shape of the graph of scale factor versus time is different from region to region.

Notice the red dashed line: for models above the line, the geometry of the universe resembles that of the surface of a sphere. Those below the line have a geometry like that of the surface of a saddle. Points on the line imply a flat universe in which the rules of Euclidean geometry apply. Figure D.41 illustrates these three models.

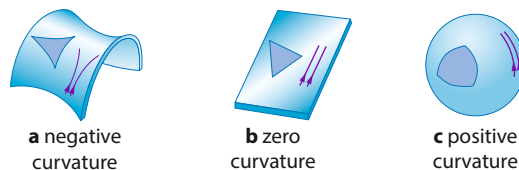


Figure D.41 Three models with different curvatures. **a** In negative-curvature models, the angles of a triangle add up to less than 180° and initially parallel lines eventually diverge. **b** Ordinary flat (Euclidean) geometry. **c** Positive curvature, in which the angles of a triangle add up to more than 180° and initially parallel lines eventually intersect. (From M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004. © The Open University, used with permission)

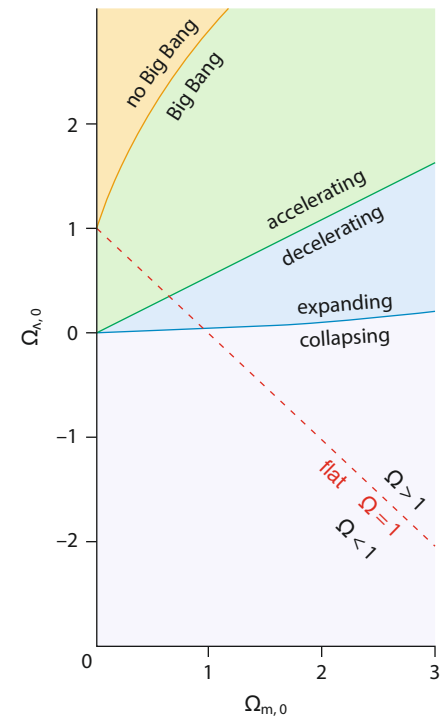


Figure D.40 There are various possibilities for the evolution of the universe depending on how much energy and mass it contains. (From M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004. © The Open University, used with permission.)

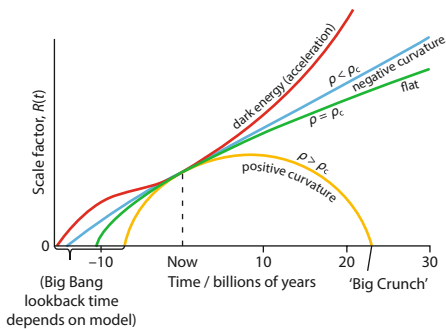


Figure D.42 Solutions of Einstein’s equations for the evolution of the scale factor. The present time is indicated by ‘now’. Notice that the estimated age of the universe depends on which solution is chosen. Three models assume zero dark energy; the one shown by the red line does not.

Of the very many possibilities, we will be interested in just four cases. The first three correspond to $\Omega_\Lambda = 0$ (they are of mainly historical interest, because observations favour $\Omega_\Lambda \neq 0$). These are shown as the orange, green and blue lines in Figure D.42.

In all three cases the scale factor starts from zero, implying a Big Bang. In one possibility (orange line), $R(t)$ starts from zero, increases



to a maximum value and then returns to zero; that is, the universe collapses after an initial period of expansion. This is called the **closed model**, and corresponds to $\Omega_m > 1$, i.e. $\rho_m > \rho_c$. The second possibility corresponds to $\Omega_m < 1$, i.e. $\rho_m < \rho_c$. Here the scale factor $R(t)$ increases without limit – the universe continues to expand forever. This is called the **open model**. The third possibility is that the universe expands forever, but with a decreasing rate of expansion, becoming zero at infinite time. This is called the **critical model** and corresponds to $\Omega_m = 1$. The density of the universe in this case is equal to the critical density: $\rho_m = \rho_c$.

Keep in mind that these three models have $\Omega_\Lambda = 0$ and so are *not* consistent with observations. The fourth case, the red line in Figure D.42, is the one that agrees with current observations. Data from the Planck satellite observatory (building on previous work by WMAP, Boomerang and COBE) indicate that $\Omega_m \approx 0.32$ and $\Omega_\Lambda \approx 0.68$. This implies that $\Omega_m + \Omega_\Lambda \approx 1$, and corresponds to the red dashed line in Figure D.42. This is consistent with the analysis of the Boomerang data that we discussed in Section D5.2, and means that at present our universe has a flat geometry and is expanding forever at an accelerating rate, and that 32% of its mass–energy content is matter and 68% is dark energy.

D5.8 Dark energy

In Section D3.5 we saw how analysis of distant Type Ia supernovae led to the conclusion that the expansion of the universe is accelerating. This ran contrary to expectations: gravity should be slowing the distant galaxies down. We also saw that an accelerating universe demands a non-zero value of the cosmological constant, which in turn implies the presence of a ‘vacuum energy’ that fills all space; this energy has been called dark energy.

The presence of this energy creates a kind of repulsive force that not only counteracts the effects of gravity on a large scale but actually dominates over it, causing acceleration in distant objects rather than the expected deceleration. The domination of the effects of dark energy over gravity appears to have started about 5 billion years ago.

There is now convincing evidence that $\Omega_m + \Omega_\Lambda \approx 1$, based on detailed studies of anisotropies in CMB radiation undertaken by COBE, WMAP, the Boomerang collaboration and the Planck satellite observatory. Data from Planck indicate that the mass–energy density of the universe consists of approximately 68% dark energy, 27% dark matter and 5% ordinary matter. This means that we understand just 5% of the mass–energy of the universe! These facts – along with the discovery (announced in March 2014 but still not confirmed) of gravitational waves, supporting another important part of the Big Bang model (inflation) – make these very exciting times for cosmology!

Nature of science

Cognitive bias

When interpreting experimental results, it is tempting to dismiss or find ways to explain away results that do not fit with the hypotheses. In the late 20th century, most scientists believed that the expansion of the universe must be slowing down because of the pull of gravity.

Evidence from the analysis of Type Ia supernovae in 1998 showed that the expansion of the universe is accelerating – a very unexpected result. Corroboration from other sources has led to the acceptance of this result, with the proposed dark energy as the cause of the acceleration.

? Test yourself

86 a Outline what is meant by the **cosmological principle**.

- b Explain how this principle may be used to deduce that the universe:
- i has no centre
 - ii has no edge.

87 State **two** pieces of observational evidence that support the cosmological principle.

88 a Outline what is meant by the **scale factor of the universe**.

- b Sketch a graph to show how the scale factor of the universe varies with time for a model with zero cosmological constant and a density greater than the critical density.
- c Draw another graph to show the variation of the CMB temperature for the model in **b**.

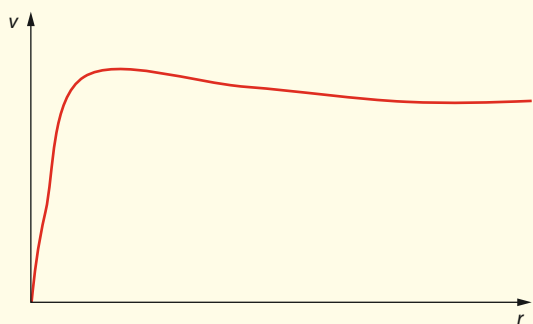
89 Sketch and label **three** graphs to show how the scale factor of the universe varies with time for models with zero cosmological constant. Use your graphs to explain why the three models imply different ages of the universe.

90 a Derive the rotation curve formula (showing the variation of speed with distance) for a mass distribution with a uniform density.

b Draw the rotation curve for **a**.

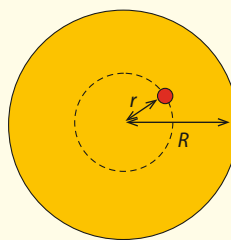
91 a Derive the rotation curve formula for a spherical distribution of mass in a galaxy that varies with the distance r from the centre according to $M(r) = kr$, where k is a constant.

- b By comparing your answer with the rotation curve below, suggest why your rotation curve formula implies the existence of matter away from the centre of the galaxy.

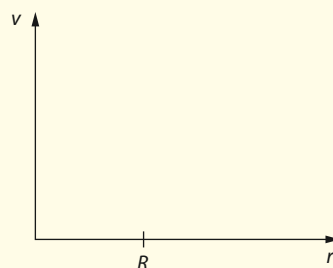


92 Sketch a graph to show the variation of the scale factor for a universe with a non-zero cosmological constant.

93 The diagram below shows a spherical cloud of radius R whose mass is distributed with constant density.



- a A particle of mass m is at distance r from the centre of the cloud. On a copy of the axes below, draw a graph to show the expected variation with r of the orbital speed v of the particle for $r < R$ and for $r > R$.



- b Describe one way in which the rotation curve of our galaxy differs from your graph.

94 a What do you understand by the term **dark matter**?

- b Give three possible examples of dark matter.

95 Distinguish between **dark matter** and **dark energy**.

96 Explain why, in an accelerating universe, distant supernovae appear dimmer than expected.

97 a Outline what is meant by the **anisotropy of the CMB**.

- b State what can be learned from studies of CMB anisotropies.

98 State how studies of CMB anisotropy lead to the conclusion that the universe is flat.



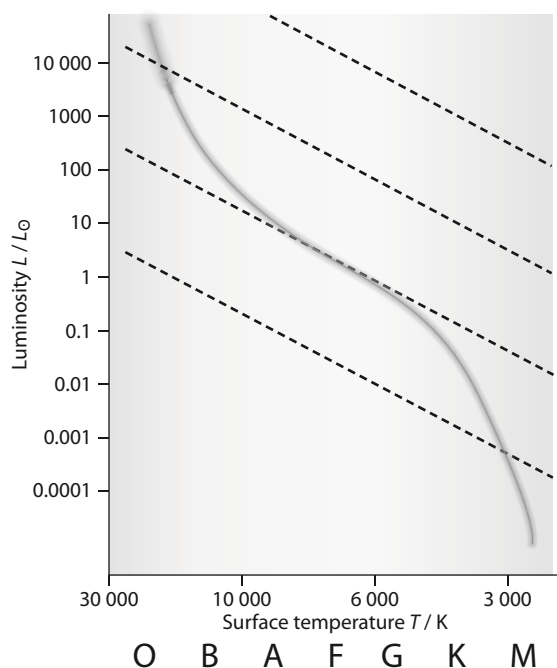
- 99 Use Figure D.40 to suggest whether current data support a model with a negative cosmological constant.
- 100 Derive the dependence $T \propto \frac{1}{R}$ of the temperature T of the CMB on the scale factor R of the universe.
- 101 a Outline what is meant by the critical density.
 b Show using Newtonian gravitation that the critical density of a cloud of dust expanding according to Hubble's law is given by
$$\rho_c = \frac{3H^2}{8\pi G}.$$

 c Current data suggest that $\Omega_m = 0.32$ and $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Calculate the matter density of the universe.
- 102 a Suggest why determination of the mass density of the universe is very difficult.
 b Estimate how many hydrogen atoms per m^3 cubic metre a critical density of $\rho_c \approx 10^{-26} \text{ kg m}^{-3}$ represents.
- 103 a State what is meant by an accelerating of the universe.
 b Draw the variation of the scale factor with time for an accelerating model.
 c Use your answer to draw the variation of temperature with time in an accelerating model.
- 104 Explain, with the use of two-dimensional examples if necessary, the terms **open** and **closed** as they refer to cosmological models. Give an example of a space that is finite without a boundary and another that is finite with a boundary.
- 105 The density parameter for dark energy is given by $\rho_\Lambda = \frac{\Lambda c^2}{3H_0^2}$. Deduce the value of the cosmological constant, given $\Omega_\Lambda = 0.68$ and $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
- 106 a List three reasons why Einstein's prediction of a constant scale factor is not correct.
 b Identify a point on the diagram in Figure D.40 where Einstein's model is located.
- 107 The Friedmann equation states that
$$H^2 = \frac{8\pi G}{3} \left(\rho + \frac{\Lambda c^2}{8\pi G} \right) - \frac{kc^2}{R^2}$$
 where the parameter k determines the curvature of the universe: $k > 0$ implies a closed universe, $k < 0$ an open universe and $k = 0$ a flat universe. Deduce the geometry of the universe given that $\Omega_m + \Omega_\Lambda = 1$.

Exam-style questions

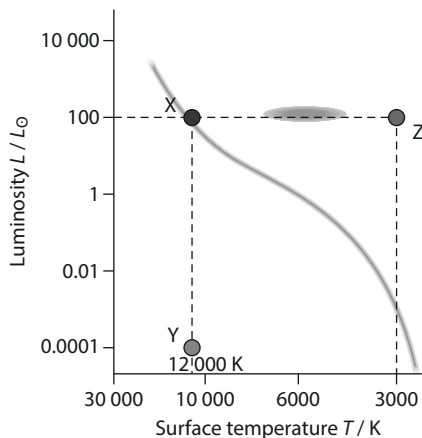
- 1 a Describe what is meant by:
- luminosity [1]
 - apparent brightness. [1]
- Achernar is a main-sequence star with a mass equal to 6.7 solar masses. Its apparent brightness is $1.7 \times 10^{-8} \text{ W m}^{-2}$ and its surface temperature is 2.6 times the Sun's temperature. The luminosity of the Sun is $3.9 \times 10^{26} \text{ W}$.
- b State **one** characteristic of main-sequence stars. [1]
- c Estimate for Achernar:
- its luminosity [2]
 - its parallax angle [2]
 - its radius in terms of the solar radius. [3]
- d
 - Describe the **method of parallax** for measuring distances to stars. [4]
 - Suggest whether the parallax method can be used for Achernar. [1]

- 2 a Suggest how the chemical composition of a star may be determined. [3]
- b Explain why stars of the following spectral classes do not show any hydrogen absorption lines in their spectra:
- i spectral class O [2]
 - ii spectral class M. [2]
- c State **one** other property of a star that can be determined from its spectrum. [1]
- 3 a Outline the mechanism by which the luminosity of Cepheid stars varies. [3]
- b Describe how Cepheid stars may be used to estimate the distance to galaxies. [4]
- c The apparent brightness of a particular Cepheid star varies from $2.4 \times 10^{-8} \text{ W m}^{-2}$ to $3.2 \times 10^{-8} \text{ W m}^{-2}$ with a period of 55 days. Determine the distance to the Cepheid. The luminosity of the Sun is $3.9 \times 10^{26} \text{ W}$. [3]
- 4 A main-sequence star has a mass equal to 20 solar masses and a radius equal to 1.2 solar radii.
- a Estimate:
- i the luminosity of this star [2]
 - ii the ratio $\frac{T}{T_{\odot}}$ of the temperature of the star to that of the Sun. [2]
- b i State **two** physical changes the star will undergo after it leaves the main sequence and before it loses any mass. [2]
- ii The star will eject mass into space in a supernova explosion. Suggest whether this will be a Type Ia or Type II supernova. [2]
- iii Describe the final equilibrium state of this star. [3]
- c On a copy of the HR diagram, draw the evolutionary path of the star. [1]





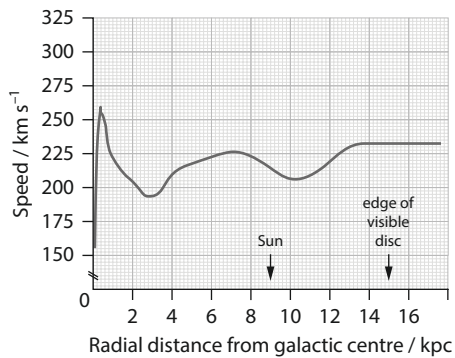
5 The HR diagram below shows three stars, X, Y and Z.



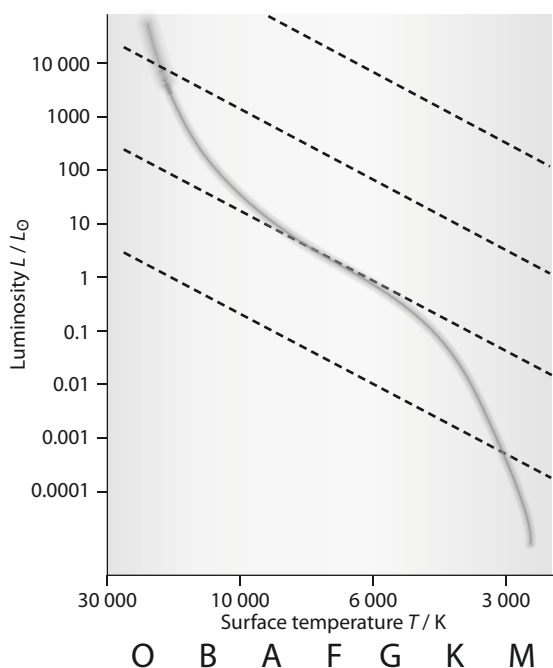
- a On a copy of the diagram, identify:
- i the main sequence [1]
 - ii the region of the white dwarfs [1]
 - iii the region of the red giants [1]
 - iv the region of the Cepheids. [1]
- b Use the diagram to estimate the following ratios of radii:
- i $\frac{R_X}{R_Y}$ [2]
 - ii $\frac{R_Z}{R_X}$. [2]
- c Estimate the mass of star X. [2]
- d
- i Show the evolutionary path of star X on the HR diagram from the main sequence until its final equilibrium state. [1]
 - ii Explain how this star remains in equilibrium in its final state. [2]
 - iii State the condition on the mass of star X in its equilibrium state. [1]
- 6 a State **Hubble's law**. [1]
- Light from distant galaxies arrives on Earth red-shifted.
- b Explain what **red-shifted** means. [2]
- c Describe the origin of this red-shift. [2]
- d Light from a distant galaxy is emitted at a wavelength of 656 nm and is observed on Earth at a wavelength of 780 nm. The distance to the galaxy is 920 Mpc.
- i Calculate the velocity of recession of this galaxy. [2]
 - ii Estimate the size of the universe when the light was emitted relative to its present size. [2]
 - iii Estimate the age of the universe based on the data of this problem. [2]
- e Outline, by reference to Type Ia supernovae, how the accelerated rate of expansion of the universe was discovered. [3]

- 7 a Outline how the CMB provides evidence for the Big Bang model of the universe. [3]
 b The photons observed today in the CMB were emitted at a time when the temperature of the universe was about 3.0×10^3 K.
 HL i Calculate the red-shift experienced by these photons from when they were emitted to the present time. [2]
 ii Estimate the size of the universe when these photons were emitted relative to the size of the universe now. [1]

- HL 8 a Show that the rotational speed v of a particle of mass m that orbits a central mass M at an orbital radius r is given by $v = \sqrt{\frac{GM}{r}}$. [1]
 b Using the result in a, show that if the mass M is instead an extended cloud of gas with a mass distribution $M = kr$, where k is a constant, then v is constant. [2]
 c The rotation curve of our galaxy is given graph below.
 i Explain how this graph may be used to deduce the existence of dark matter. [3]
 ii State **two** candidates for dark matter. [2]



- HL 9 a Describe what is meant by the **Jeans criterion**.
 b On a copy of the HR diagram below, draw the path of a protostar of mass equal to one solar mass.





- c The Jeans criterion may be expressed mathematically as $\frac{GM^2}{R} \approx \frac{3}{2}NkT$, where M and R are the mass and radius of a dust cloud, N is the number of particles in the cloud and T is its temperature.
- i Show that this is equivalent to the condition

$$R^2 \approx \frac{9kT}{8\pi G\rho m}$$

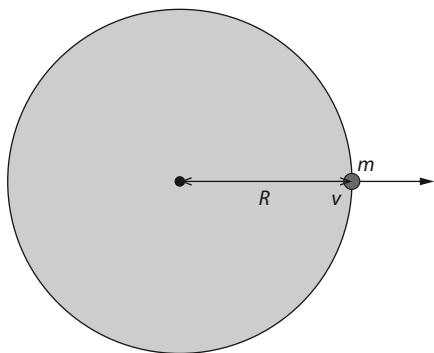
where m is the mass of a particle in the cloud of dust and ρ is the density of the cloud.

- ii Estimate the linear size R of a cloud that can collapse to form a protostar, assuming $T = 100$ K, $\rho = 1.8 \times 10^{-19} \text{ kg m}^{-3}$ and $m = 2.0 \times 10^{-27} \text{ kg}$.

[3]

[2]

- HL** 10 a The diagram below shows a particle of mass m and a spherical cloud of density ρ and radius R .



- i State the speed of the particle relative to an observer at the centre of the cloud, assuming that Hubble's law applies to this particle.

[1]

- ii Show that the total energy of the particle–cloud system is

$$E = \frac{1}{2}mR^2 \left(H_0^2 - \frac{8\pi\rho G}{3} \right)$$

[2]

- iii Hence deduce that the minimum density of the cloud for which the particle can escape to infinity is

$$\rho = \frac{3H_0^2}{8\pi G}$$

[2]

- iv Evaluate this density using $H_0 = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

- b The density found in a iii is known as the critical density. In the context of cosmological models and by reference to flat models of the universe, outline the significance of the critical density.

[2]

- c Current data suggest that the density of matter in our universe is 32% of the critical density.

- i Calculate the matter density in our universe.

[2]

- ii For this value of the matter density, draw a sketch graph to show the variation with time of the scale factor of the universe for a model with no dark energy, and also for a model with dark energy.

[2]

- HL** 11 a Outline what is meant by **fluctuations in the CMB**.

[2]

- b Give **two** reasons why these fluctuations are significant.

[4]

- HL** 12 a Explain why only elements up to iron are produced in the cores of stars.

[2]

- b Outline how elements heavier than iron are produced in the course of stellar evolution.

[4]

- c Suggest why the CNO cycle takes place only in massive main–sequence stars.

[2]

Self-test questions

Option A (HL)

1 Which list gives the postulates of special relativity?

	Postulate 1	Postulate 2
A	Moving clocks are slow.	Moving lengths are shorter.
B	The speed of light in vacuum is constant in all inertial reference frames.	The laws of physics are the same in all inertial frames.
C	It takes infinite energy to accelerate a body to the speed of light.	Moving clocks are slow.
D	Moving lengths are shorter.	The speed of light cannot be exceeded.

2 A **proper time interval** is the time between two events:

- A in the reference frame that is at rest
- B in the reference frame that moves
- C at the same point in space
- D that is the correct time interval

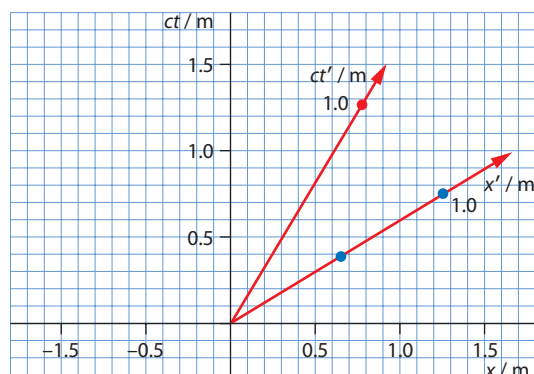
3 A rod of proper length 60 m moves past an observer with speed $0.80c$. The gamma factor for this speed is $\frac{5}{3}$. What is the length of the rod as measured by this observer?

- A 36 m
- B 48 m
- C 60 m
- D 100 m

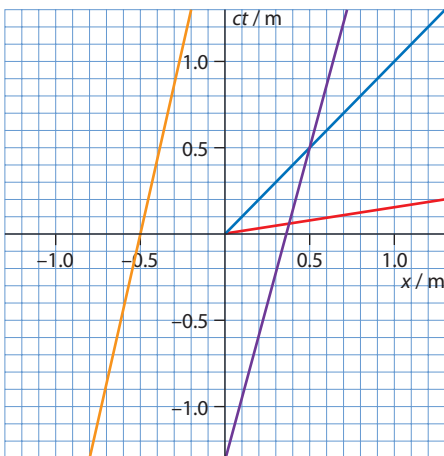
4 The figure below shows a space–time diagram for frame S (black axes) and frame S' (red axes). Two events are marked by blue dots.

The events are:

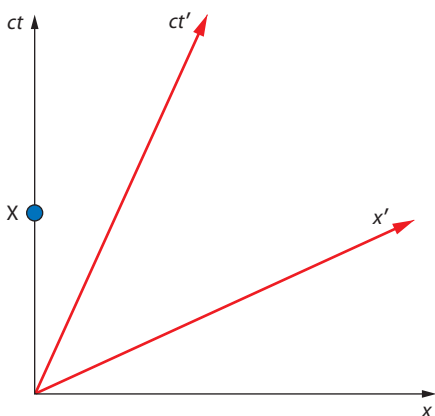
- A simultaneous in S
- B simultaneous in S'
- C simultaneous in both S and S'
- D simultaneous in neither S nor S'



- 5 The space–time diagram below shows four world lines. Which one is impossible?
A Orange
B Blue
C Red
D Purple



- 6 In the space–time diagram below, the red frame moves with velocity v past the black frame. The gamma factor for this speed is γ . In the black frame the coordinates of event X are $(x = 0, ct = 1)$. What are the coordinates of X in the red frame?

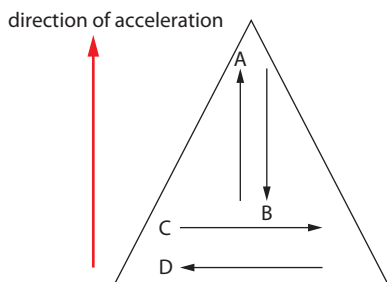


- A** $(x' = 0, ct' = \gamma)$
B $(x' = -\gamma v, ct' = \gamma)$
C $(x' = -\frac{\gamma v}{c}, ct' = \gamma)$
D $(x' = -\frac{\gamma v}{c}, ct' = 1)$
- 7 A particle has a total energy that is twice its rest energy. What is the speed of the particle?
A $\frac{\sqrt{3}c}{2}$
B $\frac{3c}{4}$
C $\frac{1c}{2}$
D $\frac{1c}{4}$

8 A particle at rest of mass M decays into two photons. What is the momentum of one of the photons?

- A $\frac{Mc}{2}$
- B Mc
- C $\frac{Mc^2}{2}$
- D Mc^2

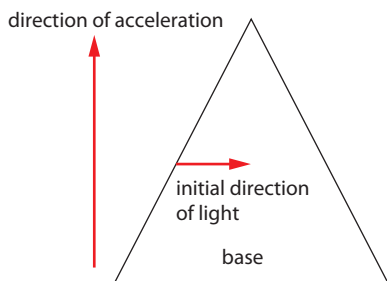
9 A frame of reference in outer space moves with constant acceleration as shown in the diagram below. Four photons are emitted. The point of emission is the beginning of the arrow and the point of reception is at the arrowhead.



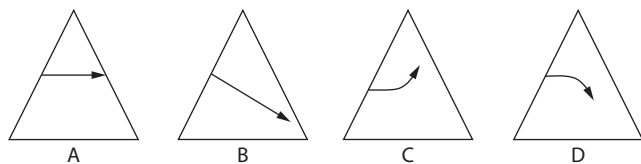
Which photon will have suffered a red-shift when it is received?

- A
- B
- C
- D

10 A frame of reference in outer space moves with constant acceleration as shown. Light is emitted in a direction that is initially parallel to the base of the frame.



What is the path of the light as observed from within the frame?



- A
- B
- C
- D

Self-test questions

Option A (SL)

1 Which list gives the postulates of special relativity?

	Postulate 1	Postulate 2
A	Moving clocks are slow.	Moving lengths are shorter.
B	The speed of light in vacuum is constant in all inertial reference frames.	The laws of physics are the same in all inertial frames.
C	It takes infinite energy to accelerate a body to the speed of light.	Moving clocks are slow.
D	Moving lengths are shorter.	The speed of light cannot be exceeded.

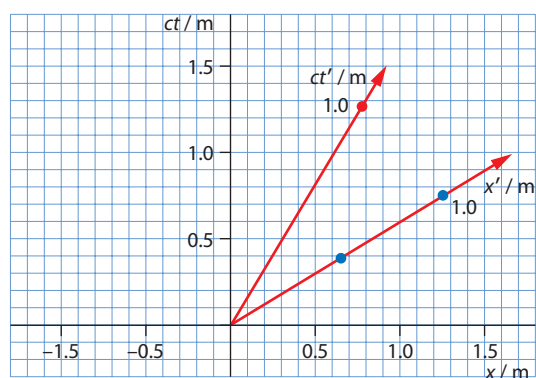
2 A **proper time interval** is the time between two events:

- A in the reference frame that is at rest
- B in the reference frame that moves
- C at the same point in space
- D that is the correct time interval

3 A rod of proper length 60 m moves past an observer with speed $0.80c$. The gamma factor for this speed is $5/3$. What is the length of the rod as measured by this observer?

- A 36 m
- B 48 m
- C 60 m
- D 100 m

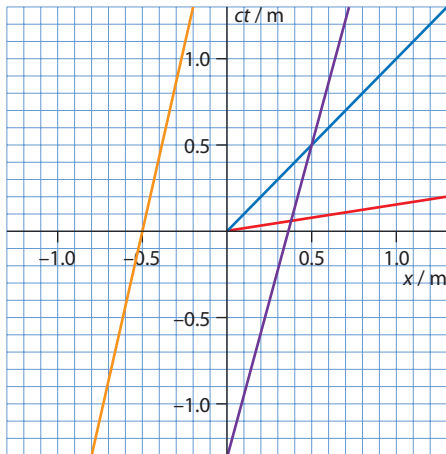
4 The figure below shows a space–time diagram for frame S (black axes) and frame S' (red axes). Two events are marked by blue dots.



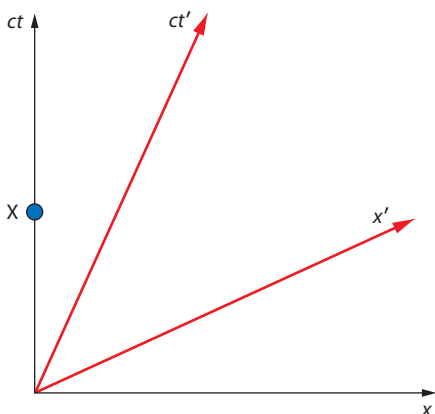
The events are:

- A simultaneous in S
- B simultaneous in S'
- C simultaneous in both S and S'
- D simultaneous in neither S nor S'

5 The space–time diagram below shows four world lines. Which one is impossible?



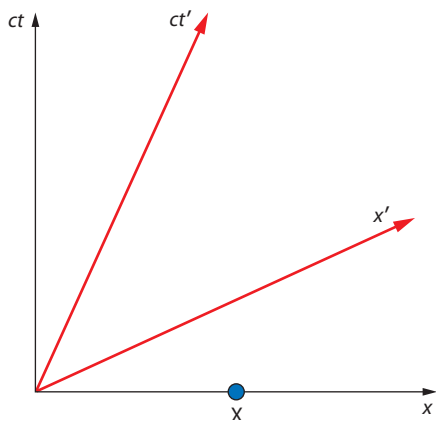
- A Orange
 B Blue
 C Red
 D Purple
- 6 In the space–time diagram below, the red frame moves with velocity v past the black frame. The gamma factor for this speed is γ . In the black frame the coordinates of event X are $(x = 0, ct = 1)$.



What are the coordinates of X in the red frame?

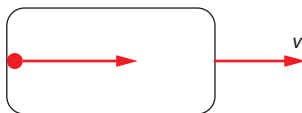
- A $(x' = 0, ct' = \gamma)$
 B $(x' = -\gamma v, ct' = \gamma)$
 C $(x' = -\frac{\gamma v}{c}, ct' = \gamma)$
 D $(x' = -\frac{\gamma v}{c}, ct' = 1)$

- 7 In the space–time diagram below, the red frame moves with velocity v past the black frame. The gamma factor for this speed is γ . In the black frame the coordinates of event X are $(x = 1, ct = 0)$.



What are the coordinates of X in the red frame?

- A $(x' = 0, ct' = \gamma)$
 B $(x' = \gamma, ct' = -\frac{\gamma v}{c})$
 C $(x' = -\frac{\gamma v}{c}, ct' = \gamma)$
 D $(x' = -\frac{\gamma v}{c}, ct' = 1)$
- 8 In frame S event 1 and event 2 happen at the same time and are separated by $\Delta x = x_2 - x_1 = L$. Which statement about the times of these vents in frame S' is correct?
- A Event 1 occurs $\frac{\gamma v L}{c}$ before event 2.
 B Event 1 occurs $\frac{\gamma v L}{c}$ after event 2.
 C Event 1 occurs $\frac{\gamma v L}{c^2}$ before event 2.
 D Event 1 occurs $\frac{\gamma v L}{c^2}$ after event 2.
- 9 A rocket moving at $0.50c$ relative to the ground launches a missile in the direction of motion of the rocket. The speed of the missile relative to the rocket is $0.50c$. What is the speed of the rocket relative to the ground?
- A c
 B $\frac{c}{1.25}$
 C $\frac{c}{0.75}$
 D $0.50c$
- 10 In the figure below, a photon leaves the left-hand wall of the box, arrives at the right-hand wall, reflects and returns to the left-hand wall. The box has proper length L and moves with velocity v to the right relative to the ground.



What is the time for total trip, as measured by an observer inside the box?

- A $\frac{L}{c}$
 B $\frac{L}{c+v} + \frac{L}{c-v}$
 C $2\frac{L}{c}$
 D $2\frac{L}{\gamma c}$

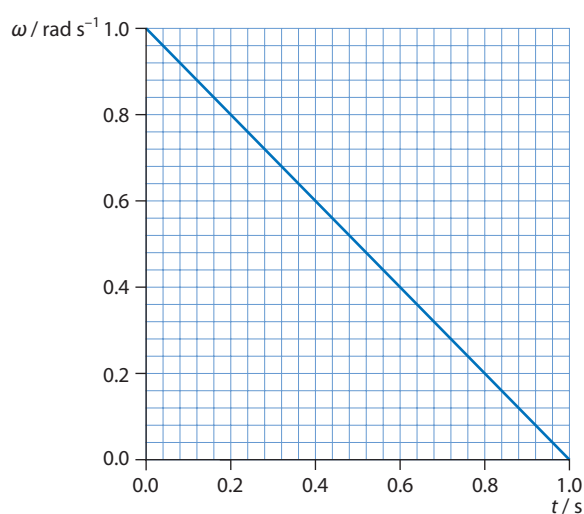
Self-test questions

Option B (HL)

1 A disc rotates about a vertical axis through its centre. The initial angular speed of the disc is 20 rad s^{-1} . It comes to rest after 40 revolutions. What is the angular deceleration of the disc?

- A $\frac{5}{2\pi} \text{ rad s}^{-2}$
- B $\frac{5}{2} \text{ rad s}^{-2}$
- C 5 rad s^{-2}
- D $\frac{5\pi}{2} \text{ rad s}^{-2}$

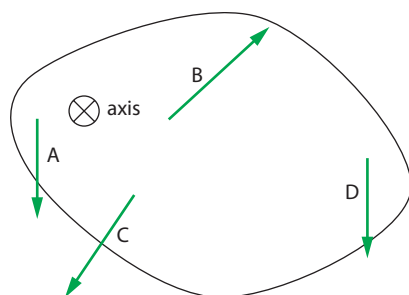
2 The graph below shows how the angular speed of a rotating body varies with time.



What do the slope and the area under the graph represent?

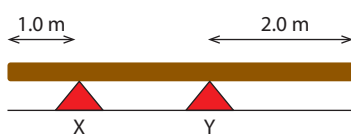
	Slope	Area
A	angular acceleration	distance travelled
B	angular acceleration	angle swept
C	linear acceleration	distance travelled
D	linear acceleration	angle swept

3 Which force has the greatest torque about the axis shown?



- A
- B
- C
- D

- 4 A uniform rod of length 5.0 m and weight 600 N is balanced on two supports as shown in the figure below.



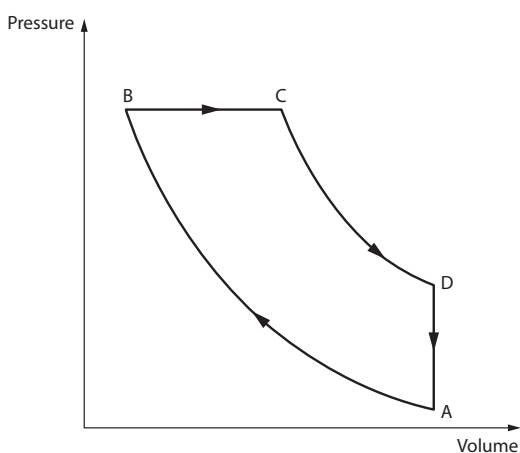
What are the forces at the two supports X and Y?

	$\frac{F_X}{N}$	$\frac{F_Y}{N}$
A	300	300
B	150	450
C	400	200
D	200	400

- 5 A sphere of mass M and radius R rolls without slipping down an inclined plane which makes an angle θ with the horizontal. (The moment of inertia of a sphere about its axis is $\frac{2}{5}MR^2$.)

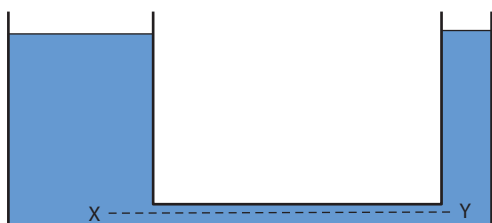
What is the linear acceleration of the sphere?

- A $g \sin \theta$
 B $\frac{2}{3}g \sin \theta$
 C $\frac{5}{7}g \sin \theta$
 D $\frac{5}{2}g \sin \theta$
- 6 A monatomic ideal gas is heated at a constant pressure of p so that the volume changes by ΔV . How much heat is provided to the gas?
- A $p\Delta V$
 B $\frac{5}{2}p\Delta V$
 C $\frac{3}{2}p\Delta V$
 D $\frac{1}{2}p\Delta V$
- 7 In the pressure–volume diagram below, AB and CD are adiabats. In which leg(s) is heat given to or taken out of the gas?



	Heat in	Heat out
A	BC and CD	DA and AB
B	DA	BC
C	DA and AB	BC and CD
D	BC	DA

- 8 An ice cube floats in a glass of water. The ice cube melts. What will happen to the level of the water in the glass?
- A It will stay the same.
 B It will decrease.
 C It will increase.
 D It will increase, decrease or stay the same, depending on the volume of water in the glass.
- 9 The figure below shows two connected columns with differing cross-sectional areas, filled with an ideal fluid. The cross-sectional area of the left-hand column is 5 times greater than the cross-sectional area of the right-hand column.



How do the pressures and densities at X and Y compare?

	Pressure	Density
A	$p_X = 5 p_Y$	$\rho_X = 5 \rho_Y$
B	$p_X = 5 p_Y$	$\rho_X = \rho_Y$
C	$p_X = p_Y$	$\rho_X = 5 \rho_Y$
D	$p_X = p_Y$	$\rho_X = \rho_Y$

- 10 A system whose natural frequency of oscillations is f_0 is subjected to damping and to an externally applied periodic force of frequency f . Under what conditions will the amplitude of oscillations be the largest?

	Damping	Frequency
A	light	$f = f_0$
B	light	$f > f_0$
C	heavy	$f = f_0$
D	heavy	$f > f_0$

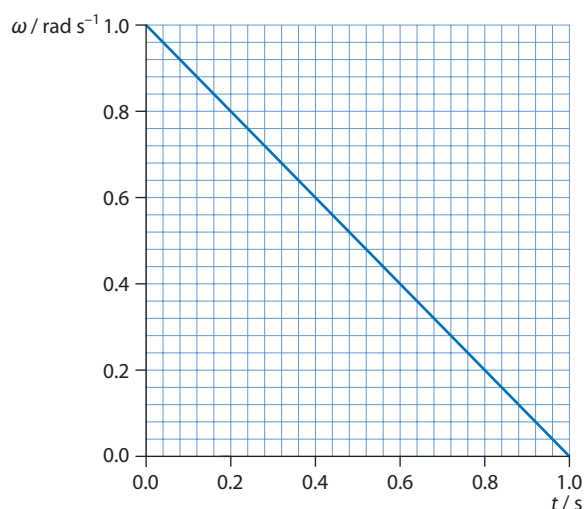
Self-test questions

Option B (SL)

1 A disc rotates about a vertical axis through its centre. The initial angular speed of the disc is 20 rad s^{-1} . It comes to rest after 40 revolutions. What is the angular deceleration of the disc?

- A $\frac{5}{2\pi} \text{ rad s}^{-2}$
- B $\frac{5}{2} \text{ rad s}^{-2}$
- C 5 rad s^{-2}
- D $\frac{5\pi}{2} \text{ rad s}^{-2}$

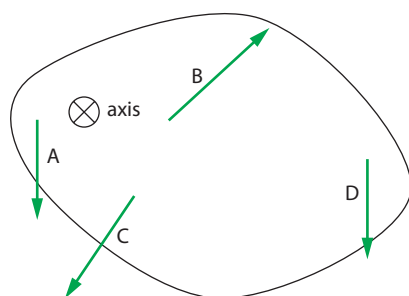
2 The graph below shows how the angular speed of a rotating body varies with time.



What do the slope and the area under the graph represent?

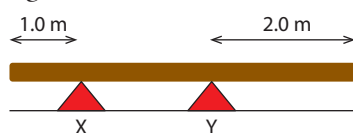
	Slope	Area
A	angular acceleration	distance travelled
B	angular acceleration	angle swept
C	linear acceleration	distance travelled
D	linear acceleration	angle swept

3 Which force has the greatest torque about the axis shown?



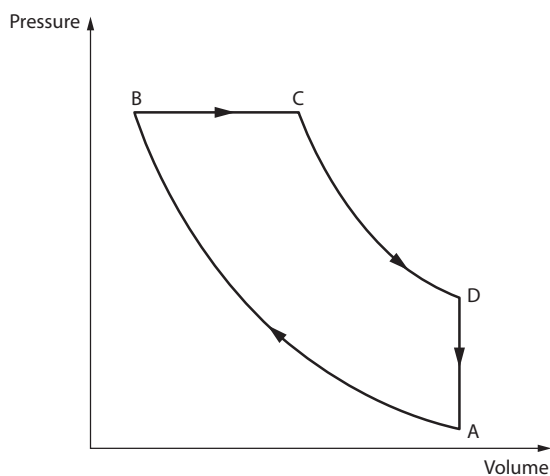
- A
- B
- C
- D

- 4 A uniform rod of length 5.0 m and weight 600 N is balanced on two supports as shown in the figure below.



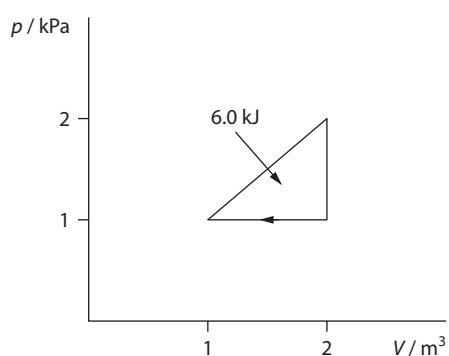
What are the forces at the two supports X and Y?

- | | $\frac{F_X}{\text{N}}$ | $\frac{F_Y}{\text{N}}$ |
|---|------------------------|------------------------|
| A | 300 | 300 |
| B | 150 | 450 |
| C | 400 | 200 |
| D | 200 | 400 |
- 5 A sphere of mass M and radius R rolls without slipping down an inclined plane which makes an angle θ with the horizontal. (The moment of inertia of a sphere about its axis is $\frac{2}{5}MR^2$.) What is the linear acceleration of the sphere?
- A $g \sin \theta$
 B $\frac{2}{3}g \sin \theta$
 C $\frac{5}{7}g \sin \theta$
 D $\frac{5}{2}g \sin \theta$
- 6 A monatomic ideal gas is heated at a constant pressure of p so that the volume changes by ΔV . How much heat is provided to the gas?
- A $p\Delta V$
 B $\frac{5}{2}p\Delta V$
 C $\frac{3}{2}p\Delta V$
 D $\frac{1}{2}p\Delta V$
- 7 In the pressure–volume diagram below, AB and CD are adiabatics. In which leg(s) is heat given to or taken out of the gas?



	Heat in	Heat out
A	BC and CD	DA and AB
B	DA	BC
C	DA and AB	BC and CD
D	BC	DA

- 8 What is the efficiency of an engine working on the cycle shown in the figure below? The heat provided to the engine is 6.0 kJ.

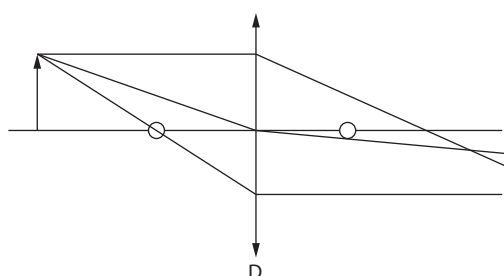
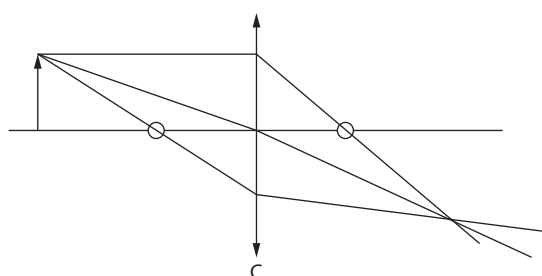
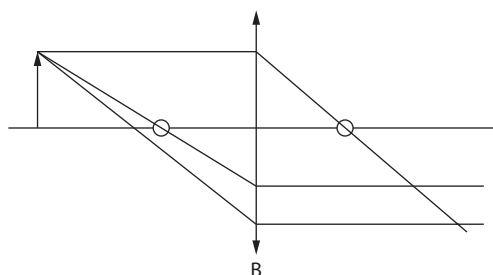
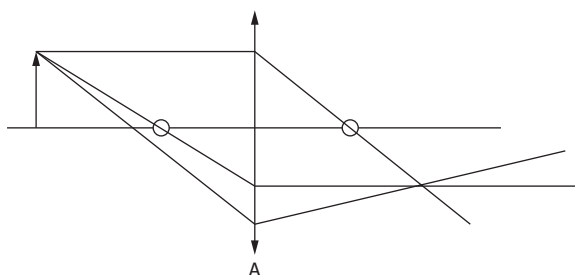


- A $\frac{1}{12}$
B $\frac{1}{4}$
C $\frac{1}{2}$
D 1
- 9 Which of the following statements is correct about an adiabatic expansion of an ideal gas?
A The internal energy remains constant.
B The pressure remains constant.
C No work is being done.
D The temperature decreases.
- 10 A quantity of heat Q is provided to an ideal gas at constant volume. The change in temperature is $\Delta\theta$. The gas does work W in expanding. How much heat must be provided to the same quantity of another ideal gas at constant pressure so that the change in temperature is $\Delta\theta$?
A Q
B $Q + W$
C $Q - W$
D W

Self-test questions

Option C (HL)

1 Of the figures below, which one is a correct ray diagram for a converging lens?



- A
- B
- C
- D

2 An object is placed at a distance of 30 cm in front of a converging lens of focal length 10 cm. What is the correct description of the image?

	Type of image	Distance
A	real	7.5 cm
B	real	15 cm
C	virtual	7.5 cm
D	virtual	15 cm

3 An object is placed at a distance of 4.0 cm in front of a converging (concave) mirror of focal length 12 cm. What is the correct description of the image?

	Type of image	Distance
A	real	3.0 cm
B	real	6.0 cm
C	virtual	3.0 cm
D	virtual	6.0 cm

- 4 Optical instrument defects include spherical and chromatic aberrations. Which list correctly identifies an instrument with an aberration?

	Spherical	Chromatic
A	lenses and mirrors	lenses and mirrors
B	lenses and mirrors	lenses
C	lenses	lenses and mirrors
D	lenses	lenses

- 5 Which list gives types of dispersion suffered by monomode and multimode optical fibres?

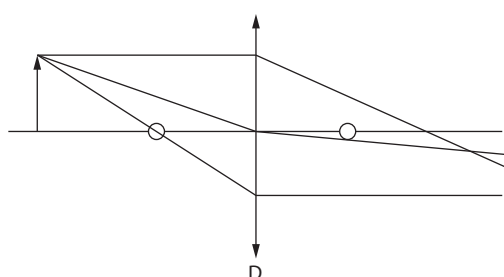
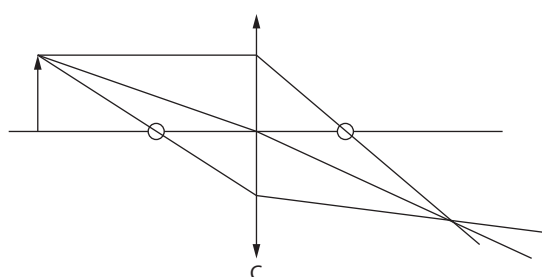
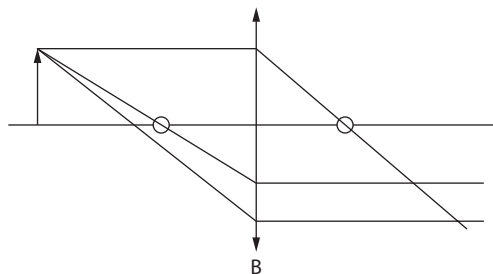
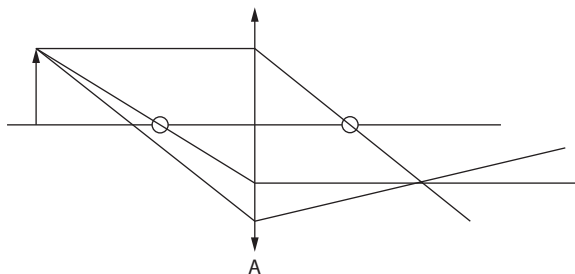
	Monomode	Multimode
A	material and waveguide	material and waveguide
B	material and waveguide	material
C	material	material and waveguide
D	waveguide	material

- 6 A digital signal of power 1200 mW is input into an optic fibre. After travelling a distance of 4.0 km in the fibre the power has been reduced to 12 mW. What is the power loss per km for this fibre?
- A 5.0 dB km^{-1}
 B 0.50 dB km^{-1}
 C 2.5 dB km^{-1}
 D 0.25 dB km^{-1}
- 7 An array of radio telescopes consists of N identical parabolic antennas of diameter D . The antennas extend over a linear distance of L . Give an **estimate** of the smallest angular separation that can be resolved.
- A $\frac{\lambda}{D}$
 B $\frac{\lambda}{ND}$
 C $\frac{\lambda}{L}$
 D $\frac{\lambda}{NL}$
- 8 In the context of X-ray imaging, **half-value thickness** is the distance in a medium after which the X-ray's:
- A wavelength is reduced by a factor of 2
 B intensity is reduced by a factor of 2
 C sharpness is reduced by a factor of 2
 D resolution is reduced by a factor of 2
- 9 In ultrasound scanning, gel is placed in between the transducer and the skin. What is the reason for this?
- A To have better contact between the transducer and the skin.
 B To make it more comfortable for the patient.
 C To make sure that no ultrasound leaks out.
 D To make sure that very little ultrasound is reflected from the skin.
- 10 In monitoring the progress of a fetus, it is advisable to use:
- A CT scan
 B X-ray scan
 C MRI
 D ultrasound

Self-test questions

Option C (SL)

1 Of the figures below, which is a correct ray diagram for a converging lens?



- A
- B
- C
- D

2 An object is placed at a distance of 30 cm in front of a converging lens of focal length 10 cm. What is the correct description of the image?

	Type of image	Distance
A	real	7.5 cm
B	real	15 cm
C	virtual	7.5 cm
D	virtual	15 cm

3 An object is placed at a distance of 4.0 cm in front of a converging (concave) mirror of focal length 12 cm. What is the correct description of the image?

	Type of image	Distance
A	real	3.0 cm
B	real	6.0 cm
C	virtual	3.0 cm
D	virtual	6.0 cm

4 Optical instrument defects include spherical and chromatic aberrations. Which list correctly identifies an instrument with an aberration?

	Spherical	Chromatic
A	lenses and mirrors	lenses and mirrors
B	lenses and mirrors	lenses
C	lenses	lenses and mirrors
D	lenses	lenses

5 Which list gives types of dispersion suffered by monomode and multimode optical fibres?

	Monomode	Multimode
A	material and waveguide	material and waveguide
B	material and waveguide	material
C	material	material and waveguide
D	waveguide	material

6 A digital signal of power 1200 mW is input into an optic fibre. After travelling a distance of 4.0 km in the fibre the power has been reduced to 12 mW. What is the power loss per km for this fibre?

- A 5.0 dB km⁻¹
- B 0.50 dB km⁻¹
- C 2.5 dB km⁻¹
- D 0.25 dB km⁻¹

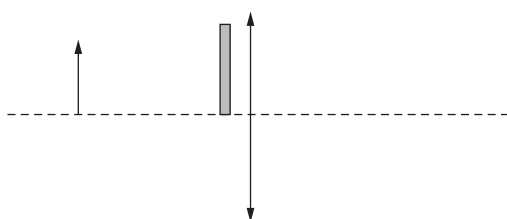
7 An array of radio telescopes consists of N identical parabolic antennas of diameter D . The antennas extend over a linear distance of L . Give an **estimate** of the smallest angular separation that can be resolved.

- A $\frac{\lambda}{D}$
- B $\frac{\lambda}{ND}$
- C $\frac{\lambda}{L}$
- D $\frac{\lambda}{NL}$

8 The final image in a refracting telescope is formed at infinity. The objective focal length is 60 cm and the eyepiece focal length is 4.0 cm. A distant object subtends an angle of 30 arcseconds at the unaided eye. What is the angle subtended at the eyepiece?

- A 2.0 arcseconds
- B 30 arcseconds
- C 450 arcseconds
- D 1800 arcseconds

9 An object is placed in front of a converging lens and real image is formed. The upper half of the lens is then covered with opaque piece of cardboard.



What, if anything, will happen to the image?

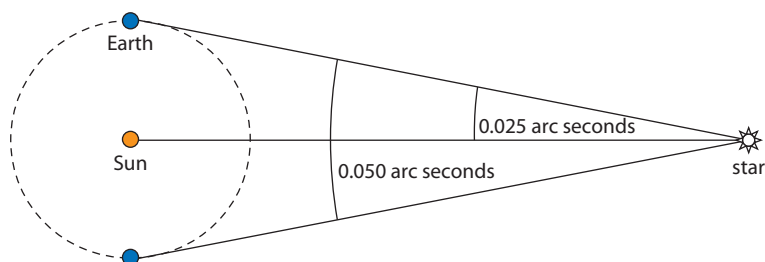
- A Nothing will happen.
- B The upper half of the image will not be formed.
- C The lower half of the image will not be formed.
- D The image will be less bright.

- 10 The angular magnification of a compound microscope is defined as the ratio of:
- A the angle subtended by the image at the eyepiece to the angle subtended by the object at the unaided eye
 - B the angle subtended by the image at the eyepiece to the angle subtended by the object at the unaided eye at a distance equal to that of the near point
 - C the angle subtended by the image at the objective to the angle subtended by the object at the unaided eye
 - D the angle subtended by the image at the objective to the angle subtended by the object at the unaided eye at a distance equal to that of the near point

Self-test questions

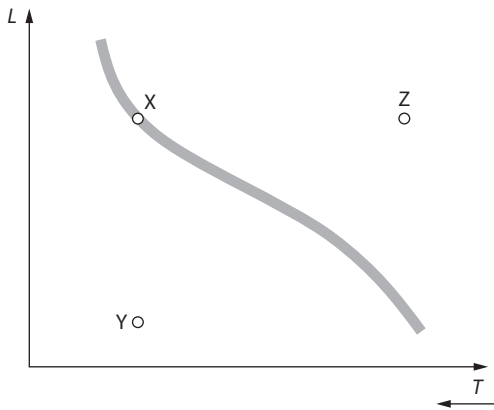
Option D (HL)

- 1 A star has double the radius and temperature of our Sun. The star's luminosity is how many times the Sun's luminosity?
A 4
B 8
C 32
D 64
- 2 A main-sequence star has double the mass of our sun. The star's luminosity is about how many times larger than that of the Sun?
A 2
B 4
C 8
D 10
- 3 The luminosity ratio of two stars X and Y is $\frac{L_X}{L_Y} = 32$ and the ratio of their apparent brightnesses is $\frac{b_X}{b_Y} = 8$. What is the value of the ratio $\frac{d_X}{d_Y}$ of their distances?
A 2
B 4
C $\frac{1}{2}$
D $\frac{1}{4}$
- 4 The diagram below shows the earth in its orbit around the Sun, and a distant star. What is the distance to the star?



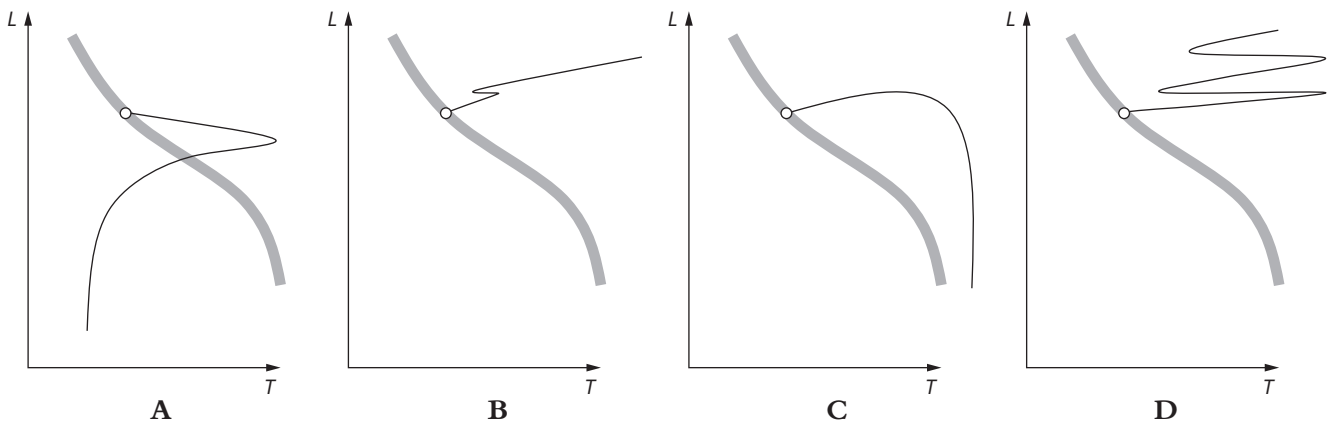
- A 40ly
- B 40pc
- C 20ly
- D 20pc

5 The HR diagram below shows three stars, X, Y and Z.



Which list gives the correct relationship between the radii of the stars?

- A $R_Y < R_X = R_Z$
 - B $R_Y = R_X < R_Z$
 - C $R_Y < R_X < R_Z$
 - D $R_Y > R_X = R_Z$
- 6 The sequence of events in the evolution of a main-sequence star of 1 solar mass include:
- A red giant \rightarrow planetary nebula \rightarrow white dwarf
 - B super red giant \rightarrow planetary nebula \rightarrow neutron star
 - C red giant \rightarrow supernova \rightarrow white dwarf
 - D super red giant \rightarrow supernova \rightarrow neutron star
- 7 On the HR diagrams below, which is the evolutionary path of a main-sequence star of 20 solar masses?



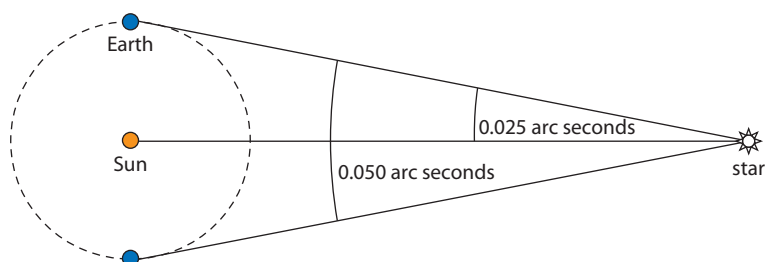
- A
- B
- C
- D

- 8 A line in the spectrum of a distant galaxy is measured to have a redshift of 0.20. The ratio of the size of the Universe when the light is received to the size of the universe when the light was emitted is:
- A 1.20
 - B $\frac{1}{1.20}$
 - C 0.20
 - D $\frac{1}{0.20}$
- 9 The existence of dark matter is inferred from:
- A fluctuations in the cosmic background radiation
 - B deviations from Hubble's law
 - C rotation curves of galaxies
 - D the fact that the universe is accelerating.
- 10 Elements with nucleon numbers above about 60 were formed:
- A during the Big Bang
 - B in fission reactions in stars
 - C in fusion reactions in stars
 - D in supernovae.

Self-test questions

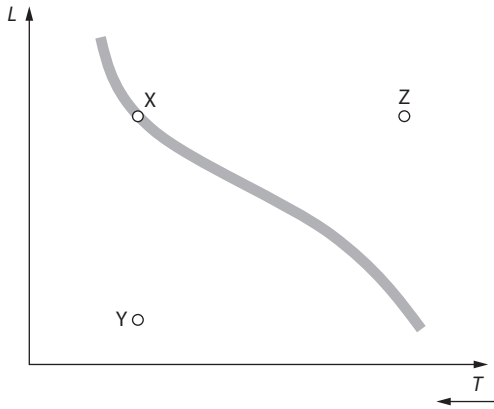
Option D (SL)

- 1 A star has double the radius and temperature of our Sun. The star's luminosity is how many times the Sun's luminosity?
A 4
B 8
C 32
D 64
- 2 A main-sequence star has double the mass of our sun. The star's luminosity is about how many times larger than that of the Sun?
A 2
B 4
C 8
D 10
- 3 The luminosity ratio of two stars X and Y is $\frac{L_X}{L_Y} = 32$ and the ratio of their apparent brightnesses is $\frac{b_X}{b_Y} = 8$. What is the value of the ratio $\frac{d_X}{d_Y}$ of their distances?
A 2
B 4
C $\frac{1}{2}$
D $\frac{1}{4}$
- 4 The diagram below shows the earth in its orbit around the Sun, and a distant star. What is the distance to the star?



- A 40ly
- B 40pc
- C 20ly
- D 20pc

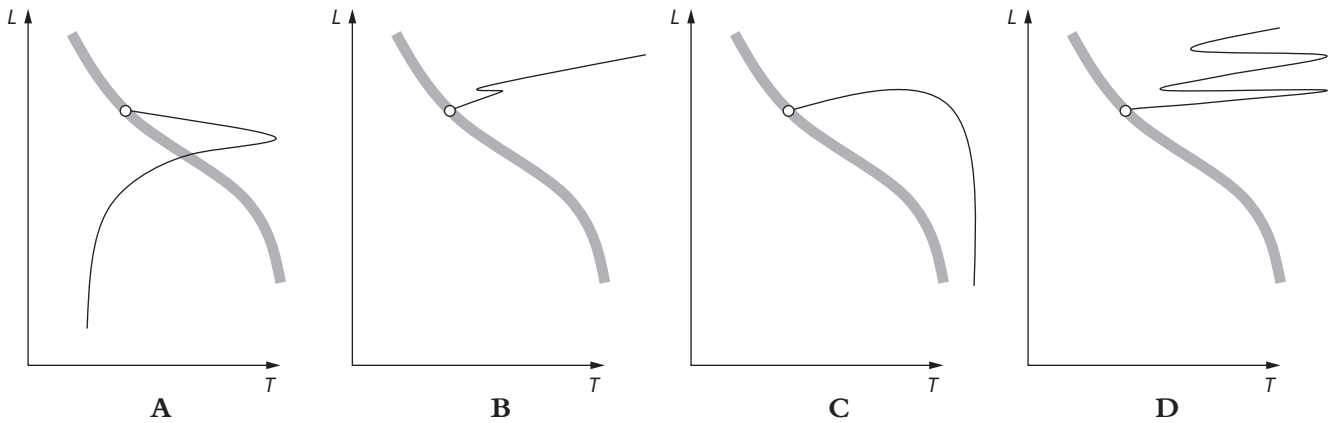
5 The HR diagram below shows three stars, X, Y and Z.



Which list gives the correct relationship between the radii of the stars?

- A $R_Y < R_X = R_Z$
 - B $R_Y = R_X < R_Z$
 - C $R_Y < R_X < R_Z$
 - D $R_Y > R_X = R_Z$
- 6 The sequence of events in the evolution of a main-sequence star of 1 solar mass include:
- A red giant \rightarrow planetary nebula \rightarrow white dwarf
 - B red supergiant \rightarrow planetary nebula \rightarrow neutron star
 - C red giant \rightarrow supernova \rightarrow white dwarf
 - D red supergiant \rightarrow supernova \rightarrow neutron star

7 Which path on the HR diagram below is the evolutionary path of a main-sequence star of 20 solar masses?



- A
- B
- C
- D

8 A line in the spectrum of a distant galaxy is measured to have a redshift of 0.20. The ratio of the size of the Universe when the light is received to the size of the universe when the light was emitted is:

- A 1.20
- B $\frac{1}{1.20}$
- C 0.20
- D $\frac{1}{0.20}$

9 The Chandrasekhar limit denotes the largest mass of:

- A a main-sequence star
- B a white dwarf
- C a red giant
- D a neutron star

10 The spectrum of a galaxy shows a redshift of z . The Hubble constant is H_0 . What is the distance to the galaxy?

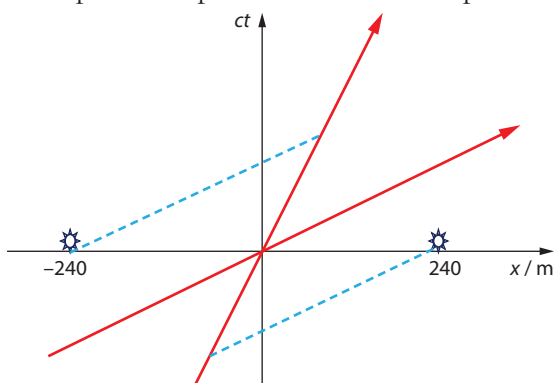
- A $\frac{cz}{H_0}$
- B $\frac{z}{cH_0}$
- C $\frac{H_0 z}{c}$
- D $\frac{cH_0}{z}$

Answers to exam-style questions

Option A

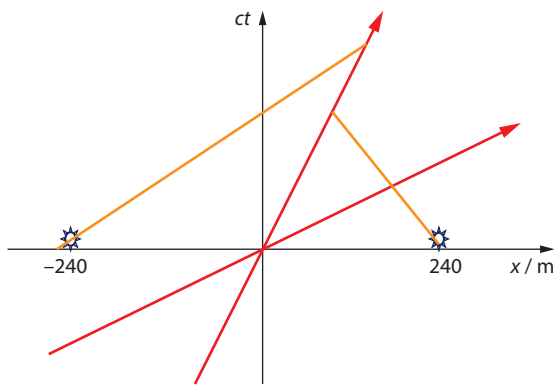
1 ✓ = 1 mark

- 1 a A reference frame is a set of rulers and clocks at every point in space. ✓
used to record the position and time of events. ✓
- b The Maxwell theory predicts that the speed of light is universal constant. ✓
Independent of the speed of the source. ✓
- c i This observer measures a non-zero magnetic field due to the current. ✓
And hence a magnetic force on the moving charged particle. ✓
- ii The observer moving along with the particle will measure an electric force on the particle. ✓
Due to the fact that the wire appears to be electrically charged. ✓
- 2 a The length of an object as measured in the rest frame of the object. ✓
- b $\gamma = \frac{1}{\sqrt{1-0.98^2}} = 5.0$ ✓
Hence the measured length is $\frac{480}{5.0} = 96$ m. ✓
- c i Light from each of the lamps will reach the observer at rest at the origin of the space station frame at the same time according to all observers. ✓
The spacecraft observers see the space station observer to move away from the light from the right lamp and towards the light from the left lamp. ✓
Light from the two lamps is moving towards the space craft observer at the same speed c . ✓
Hence for the light to arrive at the same time the light from the right must have been emitted first. ✓
- ii $\Delta x = 480$ m and $\Delta t = 0$ ✓
Hence for the space craft observers $\Delta t' = \gamma \left(\Delta t - \frac{v \Delta x}{c^2} \right) = 5.0 \times \left(0 - \frac{0.98c \times 480}{c^2} \right) = -7.8 \times 10^{-6}$ s. ✓
 $\Delta t' = -7.8 \times 10^{-6}$ s ✓
- d Lines parallel to primed x axis from lamps as shown below. ✓



- e i Line at -45° . ✓
Intersecting the primed t axis. ✓

- ii Line at 45° ✓
Intersecting the primed t axis. ✓



3 a i
$$u = \frac{0.78c + 0.50c}{1 + 0.78 \times 0.50} \checkmark$$

$$u = +0.92c \checkmark$$

ii
$$u = \frac{0.78c + (-0.50c)}{1 + 0.78 \times (-0.50)} \checkmark$$

$$u = +0.46c \checkmark$$

iii
$$u' = \frac{v - u}{1 - \frac{vu}{c^2}} = \frac{0.50c - (-0.50c)}{1 - 0.50 \times (-0.50)} \checkmark$$

$$u = +0.80c \checkmark$$

b
$$u = \frac{v + c}{1 + \frac{v}{c}} \checkmark$$

$$u = c \checkmark$$

- 4 a The twin paradox refers to a situation where two twins which are initially at the same place are separated when one of the twins flies away in a rocket and returns some time later. The “paradox” lies in that the Earth bound twin claims she is at rest and so her brother is younger when he returns. ✓
Whereas the brother may claim that in reality it is he who is at rest and he is older when the twins are reunited. ✓
- b The “paradox” is resolved by noticing that the Earth bound twin is always in the same inertial frame. ✓
Whereas the traveller changes frames and so is younger. ✓
- c i This is the time according to Earth clocks when the rocket gets to the planet, i.e. the reading is $\frac{12 \text{ ly}}{0.80c} = 15 \text{ year.} \checkmark$
- ii This is the reading of the rocket clock when the rocket gets to the planet so it is the proper time interval for the duration of the trip. ✓
Hence the reading is $\frac{15}{\gamma} = \frac{15}{5/3} = 9.0 \text{ year.} \checkmark$
- iii The rocket clock reading of 9.0 yr is the time dilated interval of the reading of the Earth clock at A. ✓
And so the reading we seek is $\frac{9.0}{\gamma} = \frac{9.0}{5/3} = 5.4 \text{ year.} \checkmark$
- iv As in iii the incoming rocket time interval from T to R. of 9.0 years is the time dilated interval from C to R $\frac{15}{\gamma} = \frac{15}{5/3} = 9.0 \text{ year.} \checkmark$
Hence the interval from C to R is 5.4 years and so the reading at C is $30 - 5.4 = 24.6 \text{ year.} \checkmark$

- v The trip took 15 year out and another 15 yr in, so the reading at R for the earth clock is 30 year. ✓
- vi For the rocket, the trip took 9 year out and another 9 yr in, so the reading at R for the incoming rocket's clock is 18 year. ✓
- d Using the previous answers the Earth bound twin aged by 30 year. ✓
And the rocket twin by 18 year. ✓
- 5 a i $\frac{690}{0.75c} = 3.1 \times 10^{-6} \text{ s}$ ✓
- ii For the space craft observers $\Delta t' = 3.1 \times 10^{-6} \text{ s}$ and $\Delta x' = 690 \text{ m}$. ✓
The gamma factor is 1.90. ✓
Hence $\Delta t = \gamma \left(\Delta t' + \frac{v \Delta x'}{c^2} \right) = 1.90 \times \left(3.1 \times 10^{-6} + \frac{0.85c \times 690}{c^2} \right) = 9.6 \times 10^{-6} \text{ s}$. ✓
- b None is proper. ✓
Because the two events happen at different points in space. ✓
- c $u = \frac{0.85c + 0.75c}{1 + 0.85 \times 0.75}$ ✓
 $u = +0.98c$ ✓
- d i $x = ut = 0.98c \times 9.6 \times 10^{-6}$ ✓
 $x = 2.8 \times 10^3 \text{ m}$ ✓
- ii For the space craft observers $\Delta t' = 3.1 \times 10^{-6} \text{ s}$ and $\Delta x' = 690 \text{ m}$ ✓
Hence $\Delta x = \gamma(\Delta x' + v\Delta t') = 1.90 \times (690 + 0.85 \times 3 \times 10^8 \times 3.1 \times 10^{-6}) = 2.8 \times 10^3 \text{ m}$. ✓
- 6 a i $\frac{6.0 \text{ ly}}{0.60c} = 10 \text{ year}$ ✓
- ii The gamma factor is 1.25. ✓
Hence the time is $\frac{10}{1.25} = 8.0 \text{ year}$. ✓
- b i The signal moves at the speed of light. ✓
 $\frac{6.0 \text{ ly}}{c} = 6.0 \text{ year}$ ✓
- ii According to the spacecraft the distance separating the earth and the craft at the time of emission of the signal is $\frac{6.0}{1.25} = 4.8 \text{ ly}$. ✓
In the time T it takes the signal to get to earth the Earth moves away distance $0.60cT$. ✓
Hence $cT = 4.8 + 0.60cT$. ✓
Giving $T = \frac{4.8 \text{ ly}}{0.40c} = 12 \text{ year}$. ✓
- 7 i For an observer on the ground, without time dilation the muons would travel a distance of $0.98c \times 2.2 \times 10^{-6} = 646.8 \approx 650 \text{ m}$ and then decay into electrons. ✓
With time dilation they would travel a distance of $0.98c \times \gamma \times 2.2 \times 10^{-6} = 5.0 \times 646.8 = 3234 \approx 3200 \text{ m}$ and then decay into electrons. ✓
The experimental fact that muons, rather than electrons, are detected at the Earth's surface is evidence for time dilation. ✓
- ii For an observer in the muon's rest frame the Earth is moving upwards and would cover a distance of $0.98c \times 2.2 \times 10^{-6} = 646.8 \approx 650 \text{ m}$ as the muons decayed into electrons. ✓
With length contraction the distance separating the surface from this observer is $\frac{3000}{\gamma} = \frac{3000}{5.0} \approx 600 \text{ m}$. ✓
Hence when the particles arrive at the surface they are muons. ✓

- 8 a i The kinetic energy gained is the work done which is 2.5 GeV. ✓
Hence the total energy is $2.5 + 0.938 = 3.4$ GeV. ✓
- ii From $E^2 = p^2c^2 + (mc^2)^2$ we find $pc = \sqrt{E^2 - (mc^2)^2} = \sqrt{3.438^2 - 0.938^2} = 3.3$ GeV. ✓
Hence $p = 3.3$ GeV c^{-1} . ✓
- iii $E = \gamma mc^2 \Rightarrow 3.4 = \gamma \times 0.938$ so that $\gamma = 3.6$. ✓
From $p = \gamma mv$ we find $3.3 \text{ GeV } c^{-2} = 3.6 \times 0.938 \text{ GeV } c^{-2} \times v$. ✓
Hence $v = \frac{3.3}{3.6 \times 0.938} c = 0.98 c$. ✓
- b The gamma factor for 0.98 c is 5.0. ✓
The total energy of one of the incoming particles is therefore 675 MeV, and the total energy is 1350 MeV. ✓
The particle is produced at rest (the initial momentum is zero) and so its rest mass is $1350 \text{ MeV } c^{-2}$. ✓
- 9 a i The energy of a particle in its rest frame/the energy measured when the particle is at rest relative to the observer. ✓
- ii The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.98^2}} = 5.0$. ✓
The total energy is $E = \gamma mc^2 = 5.0 \times 135 = 675$ MeV. ✓
The momentum is $pc = \sqrt{E^2 - (mc^2)^2} = \sqrt{675^2 - 135^2} = 661$ MeV, hence $p = 661 \text{ MeV } c^{-1}$. ✓
- b i Let p_F and p_B be the momentum of the forward and backward moving photons respectively. Then the energies of these photons are p_Fc and p_Bc . ✓
Conservation of energy and momentum implies $p_F - p_B = 661$ and $p_F + p_B = 675$. ✓
Adding gives $p_F = 668 \text{ MeV } c^{-1}$, i.e. an energy of 668 MeV. ✓
- ii From the first equation $668 - p_B = 661$. ✓
Hence $p_B = 7.0 \text{ MeV } c^{-1}$. ✓
- c i From $E = \gamma mc^2$ and $p = \gamma mv$. ✓
Eliminating the gamma factor gives the answer. ✓
- ii A photon has zero rest mass and hence $E = pc$. ✓
Substituting in $v = \frac{pc^2}{E}$ gives $v = \frac{pc^2}{pc} = c$. ✓
- 10 a A frame of reference at rest in a gravitational field is equivalent to an accelerating frame in outer space far from all masses. ✓
If the acceleration is equal to the gravitational field strength on the planet. ✓
OR
A frame of reference that is in free fall near a massive body. ✓
Is equivalent to an inertial frame of reference in outer space far from all masses. ✓
- b i Here we use the second version of the equivalence principle: consider a spacecraft in orbit around a planet and light signal sent from the back to the front of the spacecraft. ✓
The spacecraft is equivalent to an inertial frame in outer space and so the light will hit exactly the front of the spacecraft. ✓
But the spacecraft is moving on a circle and so the light signal has to bend towards the planet if it is to hit the front of the spacecraft, hence the light has bent. ✓
- ii Light rays follow geodesics, i.e. paths of least length. ✓
In a curved space these paths appear curved. ✓
- c Light rays from the quasar bend as they go past the massive galaxy. ✓
The rays arriving at the observer on earth create multiple images when extended backwards. ✓
- 11 a Gravitational red-shift is the reduction of the frequency of a photon. ✓
As it climbs higher in a gravitational field. ✓

- b** Consider a beam of light that is emitted from the base of a box on the surface of a planet ✓
 This box is equivalent to an identical box accelerating in outer space. ✓
 In this box the ray of light received at the top of the box will have its frequency reduced because of the Doppler effect since the observer is moving away from the source. ✓
 And therefore the same phenomenon will be observed in the box on the surface of the planet. ✓

c i $\frac{\Delta f}{f} = \frac{gH}{c^2} = \frac{9.8 \times 23}{90 \times 10^{16}} = 2.5 \times 10^{-15}$ ✓

The sign of $\frac{\Delta f}{f}$ is positive. ✓

- ii** The fractional change is very small. ✓

And so the emitted and received frequencies must be measured with very high precision. ✓

- iii** The frequency is the inverse of the period which may be thought to be the length in between the ticks of a clock. ✓

And since frequency changes the rate of ticking of the clock also changes. ✓

- d** In an inertial frame of reference the emitted and received frequencies will be the same. ✓

The falling elevator is equivalent to an inertial frame of reference in outer space. ✓

And so the frequency emitted and the frequency received in the elevator will also be the same. ✓

- 12 a i** A black hole is point at which the curvature of space is infinite. The falling elevator is equivalent to an inertial frame of reference in outer space. ✓

- ii** The event horizon is the set of points around the black hole where the escape speed is equal to the speed of light. ✓

b $R = \frac{2GM}{c^2} = \frac{2 \times 6.67 \times 10^{-11} \times 5.0 \times 10^{35}}{9.0 \times 10^{16}}$ ✓

$R = 7.4 \times 10^8$ m ✓

- c** The black hole constantly attracts mass from surrounding bodies into it. ✓

And as the mass increases the radius increases as well. ✓

d i From $\Delta t = \frac{\Delta t_0}{\sqrt{1 - \frac{R}{r}}}$ we find $15 = \frac{5.0}{\sqrt{1 - \frac{R}{r}}}$, i.e. $1 - \frac{R}{r} = \frac{1}{9.0}$. ✓

And so $r = 1.125R = 8.3 \times 10^8$ m. ✓

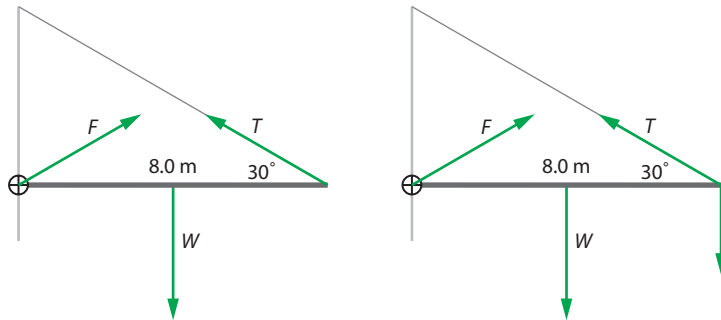
- ii** The signal will be red-shifted. ✓

Answers to exam-style questions

Option B

1 ✓ = 1 mark

1 a i The forces are as shown in the left diagram:



The perpendicular distance between the axis and the line of the tension force is $L \sin 30^\circ$. ✓
Rotational equilibrium by taking torques about an axis through the point of support gives:

$$W \times \frac{L}{2} = TL \sin 30^\circ \quad \checkmark$$

Hence $W = T = 15 \text{ kN}$. ✓

ii Translational equilibrium gives: $T \cos 30^\circ = F_x$ and $T \sin 30^\circ + F_y = 15 \text{ kN}$. ✓

Hence $T \cos 30^\circ = F_x = 12.99 \approx 13 \text{ kN}$ and $F_y = 7.5 \text{ kN}$ so that the magnitude of F is

$$F = \sqrt{12.99^2 + 7.5^2} = 15 \text{ kN}. \quad \checkmark$$

And the direction to the horizontal is $\theta = \tan^{-1} \frac{7.5}{12.99} = 30^\circ$. ✓

b The critical case is when the worker stands all the way to the right. ✓

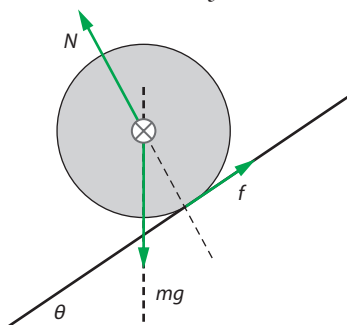
Rotational equilibrium in this case gives: $W \times \frac{L}{2} + mgL = TL \sin 30^\circ$. ✓

Solving for the tension gives: $T = 16.7 \approx 17 \text{ kN}$. ✓

2 a The forces are shown in the diagram and they are the weight of the cylinder, mg . ✓

The normal reaction, N . ✓

A frictional force f . ✓



b i Newton's second law for the translational motion down the plane is $Mg \sin \theta - f = Ma$. ✓

For the rotational motion by taking torques about the axis through the centre of mass is

$$fR = \left(\frac{1}{2} MR^2 \right) \alpha = \frac{1}{2} MRa \quad \checkmark$$

$$Mg \sin \theta - \frac{1}{2} Ma = Ma \quad \checkmark$$

From which the result follows.

ii $f = Mg \sin \theta - Ma = 12 \times 9.8 \times \sin 30^\circ - 12 \times \frac{2}{3} \times 9.8 \times \sin 30^\circ = 19.6 \approx 20 \text{ N}$ ✓

c The rate of change of the angular momentum is the net torque. ✓
And this is $fR = 19.6 \times 0.20 = 3.92 \approx 4.0 \text{ Nm}$. ✓

3 a i When the ring makes contact with the disc and while it is sliding, it exerts a frictional force on the disc but the disc exerts equal and opposite force on the ring. ✓

Hence the net torque is zero and hence angular momentum is conserved. ✓

ii The initial angular momentum of the disc is $L = I\omega = \frac{1}{2}MR^2\omega = \frac{1}{2} \times 4.00 \times 0.300^2 \times 42.0 = 7.56 \text{ Js}$. ✓

After the ring lands the total angular momentum is $L = \frac{1}{2} \times 4.00 \times 0.300^2 \times \omega + 2.00 \times 0.300^2 \times \omega$. ✓

Hence $\frac{1}{2} \times 4.00 \times 0.300^2 \times \omega + 2.00 \times 0.300^2 \times \omega = 7.56 \text{ Js}$ which gives $\omega = 21 \text{ rad s}^{-1}$. ✓

iii The initial kinetic energy is $E_K = \frac{1}{2}I\omega^2 = \frac{1}{2} \times \left(\frac{1}{2} \times 4.00 \times 0.300^2\right) \times 42.0^2 = 158.76 \text{ J}$. ✓

The final is $E_K = \frac{1}{2} \times \left(\frac{1}{2} \times 4.00 \times 0.300^2\right) \times 21.0^2 + \frac{1}{2} \times (2.00 \times 0.300^2) \times 21.0^2 = 79.38 \text{ J}$ leading to a loss of $79.38 \approx 79.4 \text{ J}$. ✓

b i $\alpha = \frac{\Delta \omega}{\Delta t} = \frac{21.0}{3.00} = 7.00 \text{ rad s}^{-2}$ ✓

ii $\theta = \frac{1}{2}\alpha t^2 = \frac{1}{2} \times 7.00 \times 3.00^2 = 31.5 \text{ rad}$ ✓

Which is $\frac{31.5}{2\pi} = 5.0$ revolutions. ✓

iii $\Gamma = \frac{\Delta L}{\Delta t}$ ✓

$\Gamma = \frac{2.00 \times 0.300^2 \times 21.0}{3.00} = 1.26 \text{ Nm}$ ✓

iv It is equal and opposite to that on the ring. ✓

Because the force on the ring is equal and opposite to that on the disc. ✓

c The change in the kinetic energy of the ring is $\frac{1}{2} \times (2.00 \times 0.300^2) \times 21.0^2 = 39.69 \text{ J}$. ✓

And so the power developed is $\frac{39.69}{3.00} = 13.2 \text{ W}$. ✓

(This can also be done through $P = \Gamma \bar{\omega} = 1.26 \times \frac{21.0}{2} = 13.2 \text{ W}$.)

4 a i The temperature at B doubled at constant volume so the pressure also doubles at $p_B = 4.00 \times 10^5 \text{ Pa}$. ✓

ii From $pV^{\frac{5}{3}} = c$ and $pV = nRT$ we find $p^{-\frac{2}{3}}T^{\frac{5}{3}} = c'$. ✓

Hence $(4.00 \times 10^5)^{-\frac{2}{3}} \times (600)^{\frac{5}{3}} = (2.00 \times 10^5)^{-\frac{2}{3}} \times T_C^{\frac{5}{3}}$ leading to $T_C = 455 \text{ K}$. ✓

iii The volume at B is $V_B = \frac{nRT_B}{p_B} = \frac{1.00 \times 8.31 \times 300}{2.00 \times 10^5} = 1.246 \times 10^{-2} \approx 1.25 \times 10^{-2} \text{ m}^3$. ✓

And so $\frac{V_B}{T_B} = \frac{V_C}{T_C} \Rightarrow V_C = V_B \frac{T_C}{T_B} = 1.246 \times 10^{-2} \times \frac{455}{300} = 1.890 \times 10^{-2} \approx 1.89 \times 10^{-2} \text{ m}^3$. ✓

b i $\Delta U_{AB} = \frac{3}{2} Rn\Delta T = +\frac{3}{2} \times 8.31 \times 1.00 \times 300 \checkmark$

$\Delta U_{AB} = +3739 \approx +3.74 \times 10^3 \text{ J} \checkmark$

ii This happens from C to A: $W = -p\Delta V = 2.00 \times 10^5 \times (1.25 - 1.89) \times 10^{-2} = -1280 \text{ J}$ and the change in internal energy is $\Delta U_{AB} = \frac{3}{2} Rn\Delta T = \frac{3}{2} \times 8.31 \times 1.00 \times (300 - 455) = -1932 \text{ J} \checkmark$

Hence $Q = \Delta U + W = -1932 - 1280 = -3212 \approx -3.21 \times 10^3 \text{ J} \checkmark$

c Any heat engine working in a cycle cannot transform all the heat into mechanical work. \checkmark

And this engine rejects heat into the surroundings as it should. \checkmark

5 a i A curve along which no heat is exchanged. \checkmark

ii An adiabatic expansion involves a piston moving outwards fast. \checkmark

Hence molecules bounce back from the piston with a reduced speed and hence lower temperature. \checkmark

b The product pressure \times volume is constant for an isothermal. \checkmark

This is the case for points A and C (product is 100 J). \checkmark

And the same is true for any other point on the curve, for example at $p = 2.00 \times 10^5 \text{ Pa}$, $V = 0.50 \times 10^{-3} \text{ m}^3$. \checkmark

c i $\frac{V_A}{T_A} = \frac{V_B}{T_B} \Rightarrow T_B = T_A \frac{V_B}{V_A} \checkmark$

$T_B = 300 \times \frac{0.38}{0.20} = 570 \text{ K} \checkmark$

At C $T_C = 300 \text{ K}$ since AC is isothermal. \checkmark

ii Using data at A: $n = \frac{pV}{RT} = \frac{5.00 \times 10^5 \times 0.20 \times 10^{-3}}{8.31 \times 300} \checkmark$

$n = 4.01 \times 10^{-2} \checkmark$

d i Energy is transferred out of the gas along C to A. \checkmark

From $Q = \Delta U + W$ and $\Delta U = 0$ we find $Q = -160 \text{ J}$. \checkmark

ii This happens from A to B: $W = 5.00 \times 10^5 \times (0.38 - 0.20) \times 10^{-3} = 90 \text{ J}$ and

$\Delta U = \frac{3}{2} \times 8.31 \times 4.01 \times 10^{-2} \times (570 - 300) = 135 \text{ J} \checkmark$

And so $Q = 135 + 90 = 225 \text{ J} \checkmark$

iii $W_{BC} = -\Delta U_{BC}$ (since BC is an adiabatic). \checkmark

And $\Delta U_{BC} = -\Delta U_{AB} = -135$ (since AC is an isothermal). \checkmark

OR

Since for the whole cycle $\Delta U = 0$, the net work is $Q_{in} - Q_{out} = 225 - 160 = 65 \text{ J}$. \checkmark

And $W_{net} = W_{AB} + W_{BC} + W_{CA} \Rightarrow 128 = 90 + W_{BC} - 160 \Rightarrow W_{BC} = 198 \text{ J} \checkmark$

iv The efficiency is $e = \frac{W_{net}}{Q_{in}} = \frac{65}{225} = 0.290$. \checkmark

6 a For an adiabatic, $pV^{\frac{5}{3}} = c$ and since $V = \frac{nRT}{p}$. \checkmark

we find $p \left(\frac{nRT}{p} \right)^{\frac{5}{3}} = c$. \checkmark

Raising to the 3rd power gives $p^3 \left(\frac{nRT}{p} \right)^5 = c^3$ and so the result. \checkmark

b i From $\frac{T^5}{p^2} = \text{constant}$ we find $\frac{320^5}{(2.0 \times 10^5)^2} = \frac{T^5}{(2.0 \times 10^6)^2}$. \checkmark

$$T = 320 \times \left(\frac{2.0 \times 10^6}{2.0 \times 10^5} \right)^{\frac{2}{3}} = 803.8 \approx 800 \text{ K} \checkmark$$

ii From $pV^{\frac{2}{3}} = \text{constant}$ we find. \checkmark

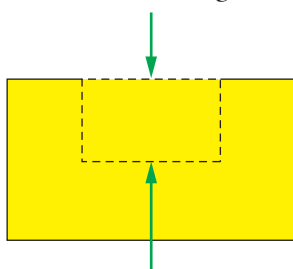
$$V = \left(\frac{2.0 \times 10^5}{2.0 \times 10^6} \right)^{\frac{3}{2}} \times 0.40 = 0.10 \text{ m}^3 \checkmark$$

c The number of moles is $n = \frac{pV}{RT} = \frac{2.0 \times 10^5 \times 0.40}{8.31 \times 320} = 30. \checkmark$

$$\Delta U = \frac{3}{2} Rn\Delta T = \frac{3}{2} \times 8.31 \times 30 \times (804 - 320) \checkmark$$

$$\Delta U = 0.181 \text{ MJ} \checkmark$$

7 Delineate a rectangular region in the liquid whose top surface is the free surface of the liquid and has area A . \checkmark



Equilibrium demands that weight = net upward force. \checkmark

In other words that $\rho Ahg = pA - p_0A$. \checkmark

From which the result follows.

b i In free fall gravity “disappears” and so the pressure is just the atmospheric pressure. \checkmark

ii When the liquid accelerates upwards there is an additional force pushing the liquid upwards and so the pressure increases. \checkmark

c A body immersed in a fluid experiences an upward force that is equal to the weight of the displaced liquid. \checkmark

d i Equilibrium demands that $\rho_{\text{wood}}Vg = \rho_{\text{water}} \times 0.75Vg$. \checkmark

$$\text{Hence } \rho_{\text{wood}} = \rho_{\text{water}} \times 0.75 = 750 \text{ kg m}^{-3} \checkmark$$

ii Equilibrium demands that $\rho_{\text{wood}}Vg = \rho_{\text{oil}} \times 0.82Vg$. \checkmark

$$\text{Hence } \rho_{\text{oil}} = \frac{\rho_{\text{wood}}}{0.82} = \frac{750}{0.82} = 914.6 \approx 910 \text{ kg m}^{-3} \checkmark$$

8 a i $p_0 + \rho gz = p_0 + \frac{1}{2} \rho v^2$ hence $v = \sqrt{2gz}$ \checkmark

$$v = \sqrt{2 \times 9.8 \times (220 + 40)} = 71.4 \approx 71 \text{ m s}^{-1} \checkmark$$

ii That the flow is laminar, \checkmark

and there are no losses of energy. \checkmark

b The flow rate is given by $Q = Av = \pi R^2 v$. \checkmark

$$\text{Hence } Q = \pi \times (0.25)^2 \times 71.4 = 14 \text{ m}^3 \text{ s}^{-1} \checkmark$$

c i $p = p_0 + \rho gh = 1.0 \times 10^5 + 1000 \times 9.8 \times 40$ \checkmark

$$p = 4.9 \times 10^5 \text{ Pa} \checkmark$$

ii The pressure is given by $p_0 + \rho gz = p + \frac{1}{2} \rho v^2$ where the speed can be found from the flow rate (i.e. the continuity equation) $\pi \times (0.65)^2 \times v = 14.02 \approx 14 \text{ m}^3 \text{ s}^{-1}$. \checkmark

i.e. $v = 10.56 \approx 11 \text{ ms}^{-1}$. ✓

And hence

$$p = p_0 + \rho gz - \frac{1}{2} \rho v^2 = 1.0 \times 10^5 + 1000 \times 9.8 \times 40 - \frac{1}{2} \times 1000 \times 10.56^2 = 4.36 \times 10^5 \approx 4.4 \times 10^5 \text{ Pa} \quad \checkmark$$

d The speed at depth h is $v = \sqrt{2gh}$. ✓

The flow rate is $Q = Av = \pi R^2 \sqrt{2gh}$ and has to equal $0.40 \text{ m}^3 \text{ s}^{-1}$. ✓

$$\text{Hence } h = \frac{1}{2g} \left(\frac{0.40}{\pi R^2} \right)^2 = \frac{1}{2 \times 9.8} \left(\frac{0.40}{\pi \times 0.03^2} \right)^2 = 7.2 \text{ m}. \quad \checkmark$$

9 a The left side is connected to the holes in the tube past which the air moves fast. ✓

Hence the pressure there is low and the liquid is higher. ✓

b Call the pressure at the top of the left column p_L and that on the right p_R . Then

$$p_L + \rho_{\text{air}} gz + \frac{1}{2} \rho v_L^2 = p_R + \rho_{\text{air}} gz + \frac{1}{2} \rho v_R^2 \quad \text{and with } v_L = 0; v_R = v, \quad \checkmark$$

$$\text{it becomes } v = \sqrt{\frac{2(p_L - p_R)}{\rho_{\text{air}}}}. \quad \checkmark$$

But $p_L - p_R = \rho gh$ which gives the result. ✓

c
$$v = \sqrt{\frac{2\rho gh}{\rho_{\text{air}}}} = \sqrt{\frac{2 \times 920 \times 9.8 \times 0.25}{1.20}}. \quad \checkmark$$

$$v = 61.3 \approx 61 \text{ m s}^{-1}. \quad \checkmark$$

10 a Smooth streamlines. ✓

Closer together above the aerofoil. ✓

b i From $p_L + \rho gz + \frac{1}{2} \rho v_L^2 = p_U + \rho gz + \frac{1}{2} \rho v_U^2$ we obtain $\Delta p = p_L - p_U = \frac{1}{2} \rho v_U^2 - \frac{1}{2} \rho v_L^2$. ✓

$$\text{Hence } F = A \Delta p = A \left(\frac{1}{2} \rho v_U^2 - \frac{1}{2} \rho v_L^2 \right) = 8.0 \times \frac{1}{2} \times 1.20 \times (85^2 - 58^2) = 18.53 \approx 19 \text{ kN}. \quad \checkmark$$

ii That the area above and below the foil are equal/that the flow is laminar. ✓

c The net upward force on the foil is about 16 kN and this is an estimate of the downward force on the fuselage. ✓
Ignoring effects of torque. ✓

d i The streamlines are no longer smooth but become eddy like and chaotic. ✓

ii Everywhere on the top side of the aerofoil and especially to the right. ✓

iii It will be drastically reduced. ✓

11 a In undamped oscillations the energy is constant and so the amplitude stays the same. ✓

In damped oscillations energy is dissipated and the amplitude keeps getting smaller. ✓

b i 8.0 s ✓

ii Correct readings of amplitudes. ✓

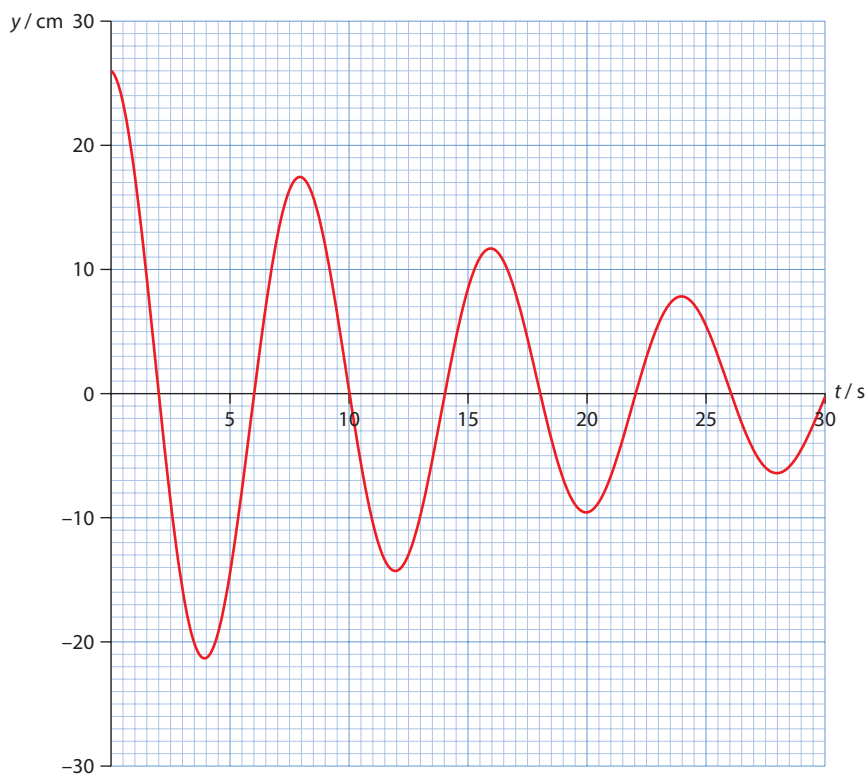
$$Q = 2\pi \frac{26^2}{26^2 - 22^2} \quad \checkmark$$

$$Q \approx 22 \quad \checkmark$$

c i Amplitude reducing more every cycle. ✓

Period staying essentially unchanged/very slightly increases. ✓

ii It will decrease. ✓



d $Q = 2\pi \frac{5.0}{5.0 - 4.6}$ ✓

$Q \approx 79$ ✓

12 a All oscillating systems have their own natural frequency of oscillation. ✓

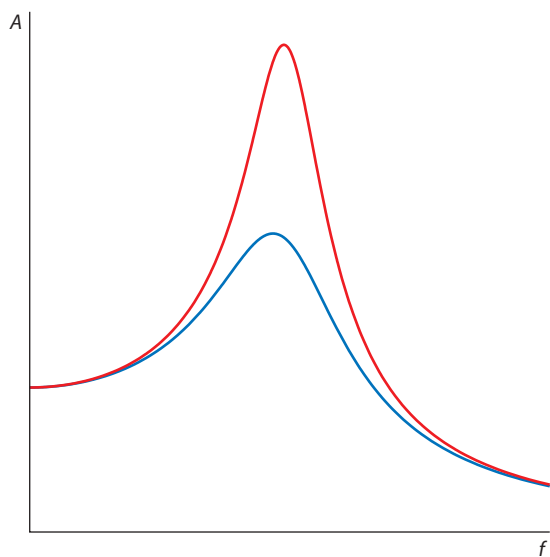
When a periodic external force is applied to the system the amplitude of oscillation will depend on the relation of the external frequency to the natural frequency. ✓

The amplitude will be large when the frequency of the external force is the same as the natural frequency in which case we have resonance external. ✓

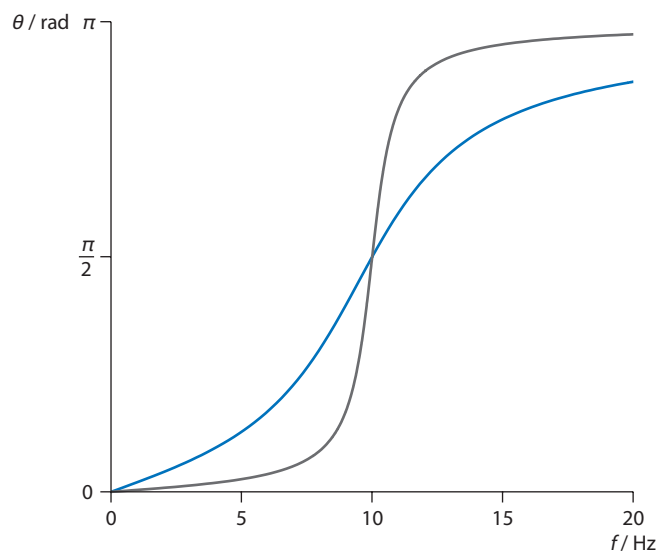
b Wider and lower curve. ✓

With peak shifted slightly to the right. ✓

See curve in blue.



- c i** Same intersection point. ✓
Less steep. ✓
See curve in blue.



- ii** 10 Hz ✓

Answers to exam-style questions

Option C

1 ✓ = 1 mark

- 1 a Rays of light parallel to the principal axis of the lens will, upon refraction through the lens, pass through the same point on the principal axis on the other side of the lens. ✓
The distance of this point from the middle of the lens is the focal length. ✓

b $M = +5.0 = -\frac{v}{u} = -\frac{v}{2.0}$ ✓

Hence $v = -10$ cm ✓

$$\frac{1}{f} = \frac{1}{2.0} + \frac{1}{-10} \Rightarrow f = 2.5 \text{ cm } v = -10 \text{ cm } \checkmark$$

c $M = -\frac{v}{u} \Rightarrow v = -Mu$ ✓

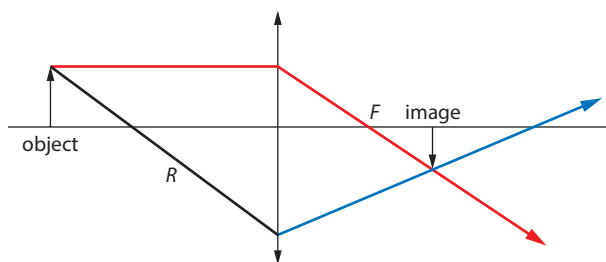
$$\frac{1}{f} = \frac{1}{u} - \frac{1}{Mu}, \text{ hence } M = \frac{f}{f - u} \checkmark$$

To make M as large as possible and still see the image the object must be placed as close to the focal point as possible and in between the focal point and the lens. ✓

- 2 a Blue line in following diagram. ✓

- b Red line in following diagram. ✓

Intersecting the principal axis at the focal point. ✓



- c It is real. ✓

Because it is formed by actual rays. ✓

- d Since half the light now goes through the lens. ✓

The image will be fully formed but not as bright as before. ✓

- e i $v = -25$ cm. ✓

$$\frac{1}{u} = \frac{1}{f} - \frac{1}{v} = \frac{1}{4.0} + \frac{1}{25} \text{ hence } u = 3.45 \approx 3.4 \text{ cm } \checkmark$$

ii $M = -\frac{v}{u} = -\frac{-25}{3.45} = +7.25$ ✓

Hence $h' = hM = 5.0 \times 7.25 = 36.2 \approx 36$ mm ✓

iii $\theta' = \tan^{-1} \frac{36.2}{250} = 8.2^\circ$ ✓

(which is much preferable to $M = 7.25 = \frac{\theta'}{\theta} = \frac{\theta'}{\frac{5}{250}} \Rightarrow \theta' = 7.25 \times \frac{5}{250} = 0.145 \text{ rad} = 8.3^\circ$)

3 a i $\frac{1}{v_1} = \frac{1}{f_o} - \frac{1}{u_1} = \frac{1}{15} - \frac{1}{20}$ ✓

Hence $v_2 = 60$ mm ✓

ii $v_2 = -250$ mm ✓

$$\frac{1}{u_2} = \frac{1}{f_c} - \frac{1}{-250} \text{ hence } u_2 = 48.4 \approx 48 \text{ mm } \checkmark$$

- b i** Angular magnification is the ratio of the angle subtended by the final image at the eyepiece. ✓
To the angle the object subtends at the unaided eye at a distance of 25 cm. ✓

ii $M = -\frac{v_1}{u_1} \times \frac{v_2}{u_2} \times \frac{D}{v_2}$ ✓

$$M = -\frac{60}{20} \times \frac{250}{48.4} = -15.5 \approx -15 \quad \checkmark$$

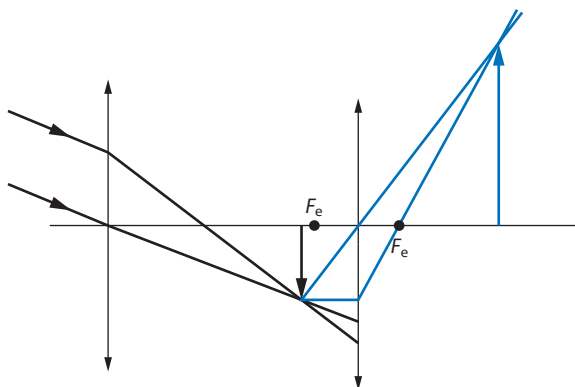
- c i** $h' = hM = 15.5 \times 8.0 = 124 \approx 120 \text{ mm}$ ✓

ii $\theta' = \tan^{-1} \frac{124}{250} = 26^\circ$ ✓

- 4 a** So that it collects a lot of light. ✓

- b** Line through middle of lens. ✓

Line parallel to PA refracting through the focal point. ✓



c $\frac{1}{u} = \frac{1}{f_e} - \frac{1}{v} = \frac{1}{0.10} - \frac{1}{0.455}$ ✓

Hence $u = 0.128 \approx 0.13 \text{ m}$ ✓

- d** $0.055 = \frac{h}{f_e} = \frac{h}{1.00} \Rightarrow h = 0.055 \text{ m}$ where h is the size of the image in the objective. ✓

The magnification of the eyepiece is $M = -\frac{0.455}{0.128} = -3.55$ and so $h' = 3.55 \times 0.055 = 0.195 \approx 0.20 \text{ m}$. ✓

- 5 a i** $\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{24} - \frac{1}{8.0}$, hence $v = -12 \text{ cm}$ ✓

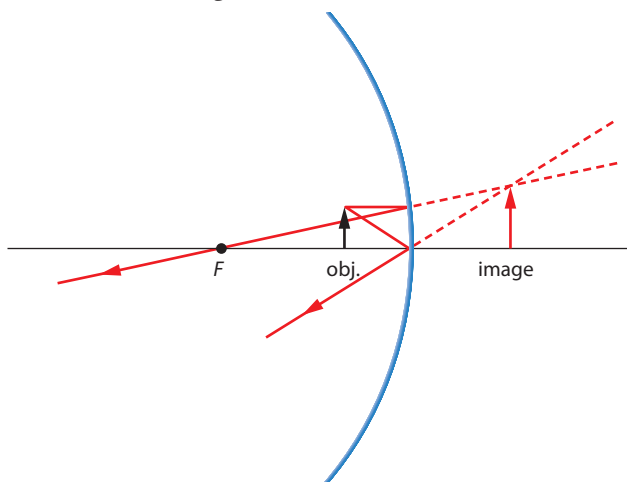
ii The magnification is $M = -\frac{-12}{8.0} = +1.5$ ✓

Hence $h' = Mh = 1.5 \times 3.0 = 4.5 \text{ cm}$ ✓

- b** One correct ray. ✓

A second correct ray. ✓

Formation of image. ✓



- c $M = -\frac{v}{u} = +0.50$, hence $v = -0.50u$. ✓
 And $u - v = 120$ cm (v is negative). ✓
 Solving for u and v we find $u = 240$ cm, $v = -120$ cm and so $f = -240$ cm (the mirror is convex). ✓
- d i Mirrors do not suffer from chromatic aberration. ✓
 So the quality of the image is better. ✓
 Large mirrors are easier to make compared to lenses. ✓
 Large mirrors are needed in order to collect the faint light of distant objects. ✓
- ii They do not suffer from spherical aberration. ✓
- 6 a $\theta = \frac{h}{f}$ hence $1.50 \times 10^{-4} = \frac{h}{10}$ ✓
 Giving $h = 1.50$ mm. ✓
- b i This image serves as a virtual object for the convex mirror and we have that $u = -1.00$ m and $v = +9.00$ m. ✓
 Hence $\frac{1}{f} = \frac{1}{u} + \frac{1}{v} = \frac{1}{-1.00} + \frac{1}{9.00}$ leading to $f = -1.125 \approx -1.12$ m. ✓
- ii $M = -\frac{v}{u} = -\frac{9.00}{-1.00} = +9.00$ ✓
- iii $h' = Mh = 9.00 \times 1.50 = 13.5$ mm ✓
- c i The image at C must be at focal point of the converging lens. ✓
 Hence $\theta \approx \frac{13.5}{120} = 0.112$ rad ✓
- ii $M = \frac{0.112}{1.5 \times 10^{-4}}$ ✓
 $M = 747 \approx 750$ ✓
- 7 a Lens aberrations denote deviations from the perfect geometrical behaviour of lenses in which the focal length is assumed to be the same for all rays and all wavelengths. ✓
- b i Spherical aberration refers to the fact that not all paraxial rays have the same focal length. ✓
 Rays far from the principal axis have a shorter focal length than rays close to the axis. ✓
 Chromatic aberration has to do with the fact that rays of different wavelength have different focal lengths. ✓
 Blue wavelengths have a shorter focal length than red. ✓
- ii Spherical aberration is reduced by not allowing rays far from the lens to enter the lens and form the image. ✓
 Chromatic aberration is corrected by using a second diverging lens of different refractive index adjacent to the first. ✓
- 8 a i $\frac{1}{v_A} = \frac{1}{f} - \frac{1}{u} = \frac{1}{24} - \frac{1}{40}$ hence $v_A = 60$ cm ✓
 $\frac{1}{v_B} = \frac{1}{f} - \frac{1}{u} = \frac{1}{24} - \frac{1}{30}$ hence $v_B = 120$ cm ✓
 Hence the difference is 60 cm. ✓
- ii $M_A = -\frac{60}{40} = -1.5$ hence $h_A = -1.5 \times 5.0 = -7.5$ cm ✓
 $M_B = -\frac{120}{30} = -4.0$ hence $h_B = -4.0 \times 5.0 = -20$ cm ✓
 Hence the difference is 12.5 cm. ✓
- b i The images of the front and the back of the rod are at 60 cm and 120 cm. ✓
 So the length of the image of the rod is different from that of the rod itself. ✓
- ii The images of the front and the back of the rod are at different heights. ✓
 And so the rod is not parallel. ✓

- 9 a i $1.58 \times \sin \theta_c = 1.45 \times \sin 90^\circ$ ✓
 $\theta_c = \sin^{-1} \frac{1.45}{1.58} = 66.595^\circ \approx 66.6^\circ$ ✓
- ii The angle of refraction at the air-core boundary must be $90^\circ - 66.595^\circ = 23.405^\circ \approx 23.4^\circ$. ✓
Hence $1.00 \times \sin A = 1.58 \times \sin 23.405^\circ$ giving $A = \sin^{-1}(1.58 \times \sin 23.405^\circ) = 38.9^\circ$. ✓
- b Material dispersion refers to the fact that rays of different wavelength have different speeds in the same medium. ✓
And so will take different times to complete a given path. ✓
Waveguide dispersion has to do with rays of light following different paths in an optic fibre and hence taking different times to arrive at their destination. ✓
- c i Waveguide dispersion may be reduced by using monomode fibres in which all rays essentially follow the same path. ✓
And by using graded index fibres in which the refractive index of the core decreases as one moves from the central axis towards the cladding. ✓
Because rays that move far from the axis now move faster (since the refractive index is less) and so cover the longer distance at higher speed essentially arriving at the same time as the rays close to the axis. ✓
- ii Material dispersion may be reduced by using monochromatic light in the transmission. ✓
- d i Scattering off impurities in the core. ✓
- ii The attenuation is $10 \log \frac{P_f}{P_i} = 10 \log \frac{25 \times 10^{-3}}{15 \times 10^{-6}}$ ✓
 $= 32.2 \text{ dB}$ ✓
Hence we need amplification after $\frac{32.2}{3.50} = 9.2 \text{ km}$. ✓
- 10 a i Ultrasound is sound of frequency higher than 20 kHz. ✓
- ii X-rays are ionising which means they do damage to cells, ultrasound does not. ✓
- b i $\frac{I_t}{I_0} = \frac{(410 - 1.8 \times 10^6)^2}{(410 + 1.8 \times 10^6)^2} \approx 1$ ✓
- ii The greatly different impedances imply that essentially all of the ultrasound gets reflected. ✓
Leaving none to be transmitted into the body for imaging. ✓
Thus a gel substance of impedance similar to tissue is placed in between the source of ultrasound and tissue. ✓
- c The pulse covers double the required distance. ✓
And so the distance is $\frac{1500 \times 6.5 \times 10^{-3}}{2} = 4.9 \text{ cm}$. ✓
- 11 a The distance that X-rays must travel through so that their intensity is reduced by a factor of 2. ✓
- b i $\mu x_{1/2} = \ln 2$ so that $\mu = \frac{\ln 2}{x_{1/2}} = \frac{\ln 2}{4.10} = 0.1691$ ✓
 $I = I_0 e^{-\mu x}$ so that $0.650 = e^{-0.1691x}$ ✓
 $\ln(0.650) = -0.1691x$ so $x = \frac{\ln(0.650)}{-0.1691} = 2.55 \text{ mm}$ ✓
- ii A larger HVT value results in smaller attenuation coefficient. ✓
And so a larger distance through which the X-rays would travel to be reduced to 65% in intensity. ✓
- c i The blurring of the image is mainly caused by X-rays scattering through the body. ✓
And may be reduced by placing metal strips in the direction of the incident X-rays so that scattered X-rays would be blocked. ✓
- ii The exposure time may be reduced by the use of intensifying screens. ✓
In these, X-rays cause the emission of visible light photons which expose photographic film faster than X-rays. ✓

- 12 Protons in the body align themselves parallel or anti-parallel to an externally supplied strong external field. ✓
A radio frequency signal forces protons to make transitions from the spin up to the spin down state. ✓
Protons make a transition to the spin up state emitting photons. ✓
The frequency of the emitted photons depends on the magnetic field at the point of emission so a second magnetic field is applied so that different parts of the body are exposed to a different net magnetic field. ✓
Knowing the net magnetic field at a given point in the body allows the frequency to be calculated and so measuring the emitted frequency is equivalent to locating the point of emission. ✓
Measuring the rate of transitions allows determination of the tissue type. ✓

Answers to exam-style questions

Option D

1 ✓ = 1 mark

- 1 a i Luminosity is the total power radiated by a star. ✓
ii Apparent brightness is the power received per unit area. ✓

b They all fuse hydrogen into helium. ✓

c i Using the mass-luminosity relation $\frac{L_A}{L_\odot} = \left(\frac{M_A}{M_\odot}\right)^{3.5}$ ✓

$$\text{Hence } L_A = L_\odot \left(\frac{M_A}{M_\odot}\right)^{3.5} = L_\odot \times 6.7^{3.5} = 778 \times 3.9 \times 10^{26} = 3.04 \times 10^{29} \approx 3.0 \times 10^{29} \text{ W } \checkmark$$

ii From $b = \frac{L}{4\pi d^2}$ we find $d = \sqrt{\frac{L}{4\pi b}} = \sqrt{\frac{3.04 \times 10^{29}}{4\pi \times 1.7 \times 10^{-8}}} = 1.19 \times 10^{18} \text{ m} = \frac{1.19 \times 10^{18}}{3.09 \times 10^{16}} \text{ pc} = 38.5 \approx 38 \text{ pc } \checkmark$

$$\text{Hence } p = \frac{1}{d} = \frac{1}{38.5} = 0.0260 \approx 0.026'' \checkmark$$

iii $\frac{L_A}{L_\odot} = 778 = \frac{\sigma 4\pi R_A^2 T_A^4}{\sigma 4\pi R_\odot^2 T_\odot^4} = \frac{R_A^2 T_A^4}{R_\odot^2 T_\odot^4} \checkmark$

$$778 = \frac{R_A^2}{R_\odot^2} \times 2.6^4 \checkmark$$

$$\text{Hence } R_A = R_\odot \frac{778}{2.6^4} \approx 17R_\odot \checkmark$$

- d i The parallax method measures the position of a star two times six months apart. ✓
The shift of the angular position of the star relative to the background of the distance stars. ✓
Allows measurement of the parallax angle which is the angle subtended by the earth's orbit radius at the star. ✓

The distance in pc is the reciprocal of the parallax angle in arc seconds. ✓

ii Yes because it is larger than the limit of 0.01 arc seconds. ✓

2 a Light reaching the Earth must go through the outer layers of the star. ✓

Photons whose energy corresponds to differences in energy between energy levels of the atoms of the star may be absorbed and so will be missing in the received light. ✓

Because the energy level differences are specific atoms determination of the chemical composition is then possible. ✓

b i Type O stars are very hot and most of the hydrogen is ionised. ✓

Hence hydrogen cannot absorb any photons. ✓

ii An M type star is relatively cool so that hydrogen atoms are mostly in their ground state. ✓

And these can only absorb ultraviolet photons not visible light photons. ✓

c The surface temperature/its magnetic field/its rotational speed. ✓

3 a The surface of the star periodically expands and contracts. ✓

It expands because radiation ionises helium atoms in the star's outer layers and the released electrons heat up the star expanding it. ✓

When most of the helium is ionised, radiation leaves the star so the star cools and contracts. ✓

b There is a relation between the average luminosity of the star and the period of variation of the luminosity. ✓

So measuring the period gives the luminosity. ✓

Measuring the (average) apparent brightness (by observing the star over a period of days) allows determination of the distance. ✓

To measure the distance to a galaxy it must be determined whether a specific Cepheid star belongs to that galaxy. ✓

- c The average apparent brightness is estimated to be $2.8 \times 10^{-8} \text{ W m}^{-2}$. ✓
 A period of 55 days corresponds to an average luminosity of about 20000 solar luminosities. ✓

$$\text{And so } d = \sqrt{\frac{L}{4\pi b}} = \sqrt{\frac{20000 \times 3.9 \times 10^{26}}{4\pi \times 2.8 \times 10^{-8}}} \approx 5 \times 10^{18} \text{ m} \checkmark$$

4 a i Using the mass-luminosity relation $\frac{L_A}{L_\odot} = \left(\frac{M_A}{M_\odot}\right)^3$. ✓

$$\text{Hence } L_A = L_\odot \left(\frac{M_A}{M_\odot}\right)^{3.5} = L_\odot \times 20^{3.5} = 3.6 \times 10^4 \times 3.9 \times 10^{26} = 1.4 \times 10^{31} \text{ W} \checkmark$$

ii $\frac{L_A}{L_\odot} = 3.6 \times 10^4 = \frac{\sigma 4\pi R^2 T^4}{\sigma 4\pi R_\odot^2 T_\odot^4} = 1.2^2 \times \frac{T^4}{T_\odot^4}$ ✓

$$\text{Hence } \frac{T}{T_\odot} = \sqrt[4]{\frac{3.6 \times 10^4}{1.2^2}} = 12.6 \approx 13 \checkmark$$

- b i The surface temperature decreases. ✓

And the radius increases. ✓

- ii A type II supernova is the explosion of a massive star after it has entered the red supergiant phase while a type Ia supernova involves mass accretion onto a white dwarf. ✓

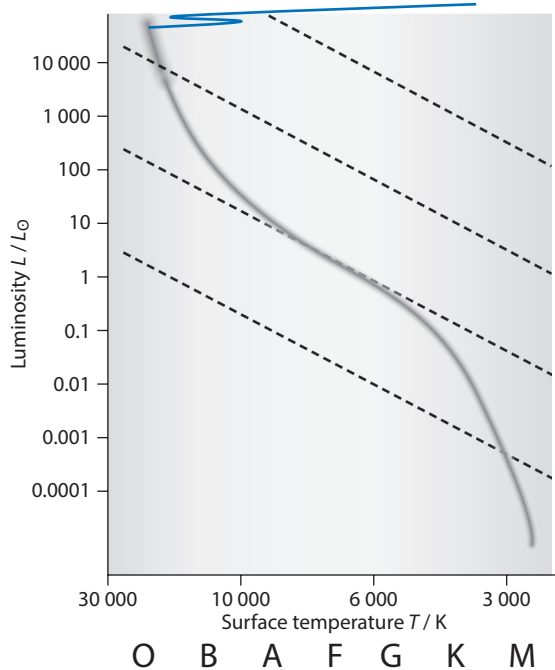
So in this case we have a type II supernova. ✓

- iii The star will be a neutron star. ✓

In which the neutron degeneracy pressure. ✓

Balances the gravitational pressure in the star. ✓

- c Blue line as shown starting approximately at the correct point. ✓



- 5 a i Strip from top left diagonally to bottom right. ✓
 ii Region in lower left. ✓
 iii Region to the right above main sequence. ✓
 iv Region joining main sequence to red giants. ✓

b i $\frac{L_X}{L_Y} = \frac{100}{0.01} = 10^4 = \frac{\sigma 4\pi R_X^2 T_X^4}{\sigma 4\pi R_Y^2 T_Y^4} = \frac{R_X^2}{R_Y^2}$ ✓

$$\text{Hence } \frac{R_X}{R_Y} = 10^2 \checkmark$$

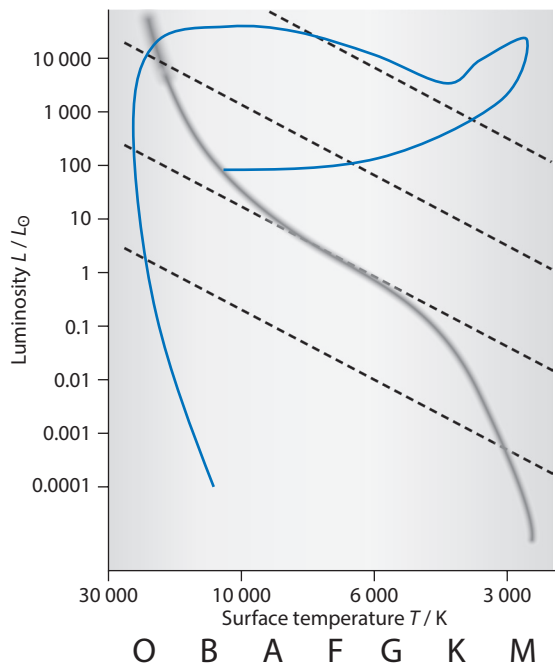
$$\text{ii } \frac{L_X}{L_Z} = 1 = \frac{\sigma 4\pi R_X^2 T_X^4}{\sigma 4\pi R_Z^2 T_Z^4} = \frac{R_X^2}{R_Z^2} \times \left(\frac{12000}{3000}\right)^4 = \frac{R_X^2}{R_Z^2} \times 256 \quad \checkmark$$

$$\text{Hence } \frac{R_X}{R_Z} = \sqrt{\frac{1}{256}} = \frac{1}{16} \quad \checkmark$$

$$\text{c } \text{Using the mass-luminosity relation } \frac{L_X}{L_\odot} = 100 = \left(\frac{M_X}{M_\odot}\right)^{3.5} \quad \checkmark$$

$$\text{Hence } \frac{M_X}{M_\odot} = 100^{1/3.5} = 3.7 \quad \checkmark$$

d i Line similar to blue line on HR diagram. \checkmark



ii Electron degeneracy pressure. \checkmark

Balances the gravitational pressure in the star. \checkmark

iii It has to be less than the Chandrasekhar limit of 1.4 solar mass. \checkmark

6 a Distant galaxies appear to move away from us with a speed that is proportional to their distance. \checkmark

b The received wavelength. \checkmark

Is larger than the wavelength at emission. \checkmark

c On a large scale, the space between galaxies stretches. \checkmark

Thus the wavelength of light emitted from a distant galaxy also stretches so it is larger at reception. \checkmark

$$\text{d i } z = \frac{\Delta\lambda}{\lambda_0} = \frac{780 - 656}{656} = 0.189 \quad \checkmark$$

$$z = \frac{v}{c} \Rightarrow v = 5.67 \times 10^7 \approx 5.7 \times 10^7 \text{ m s}^{-1} \quad \checkmark$$

$$\text{ii } z = \frac{R}{R_0} - 1 = 0.189, \text{ i.e. } \frac{R}{R_0} = 1.189 \quad \checkmark$$

$$\frac{R_0}{R} = 0.84 \quad \checkmark$$

iii The data give a Hubble value constant of: $v = Hd \Rightarrow H = \frac{5.67 \times 10^7}{920 \times 10^6 \times 3.09 \times 10^{16}} = 1.99 \times 10^{-18} \text{ s}^{-1} \quad \checkmark$

$$\text{Hence } T = \frac{1}{H} = \frac{1}{1.99 \times 10^{-18}} = 5.0 \times 10^{17} \text{ s } \left(= \frac{5.0 \times 10^{17}}{365 \times 24 \times 3600} = 1.6 \times 10^{10} \text{ year} \right) \quad \checkmark$$

e To see whether the universe accelerates or decelerates in its expansion distant objects of large redshift had to be investigated. ✓

In order to establish the relation between distance and redshift. ✓

Type Ia supernovae were chosen because their peak luminosity is known and hence the distance could be established by measuring the apparent brightness. ✓

7 a According to the hot big bang model the early universe contained radiation at very high temperature. ✓
As the universe expanded it cooled and the peak wavelength shifted to large microwave wavelengths with a black body spectrum. ✓

Which is what is being observed. ✓

b i $z = \frac{R}{R_0} - 1 = \frac{T_0}{T} - 1$ ✓

$z = \frac{3 \times 10^3}{2.7} - 1 = 1110$ ✓

ii $\frac{R}{R_0} - 1 = 1110$ ✓

Hence $\frac{R_0}{R} = 9 \times 10^{-4}$ ✓

8 a The result follows from $\frac{GMm}{r^2} = m \frac{v^2}{r}$. ✓

b With $M = kr$ the result in a becomes $v = \sqrt{\frac{Gkr}{r}} = \sqrt{Gk}$ a constant. ✓

c i The rotation curve becomes flat at large distances from the galactic centre. ✓

This is consistent with a mass distribution as in b. ✓

In other words with substantial mass far from the galactic centre. ✓

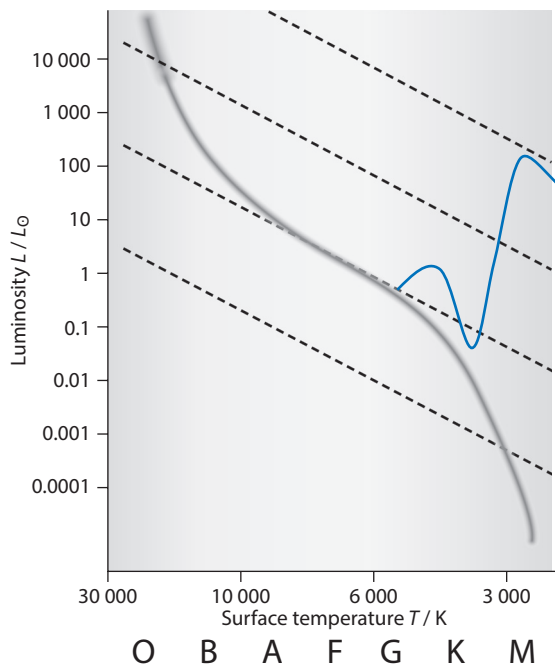
ii Small planets/brown dwarfs/black dwarfs. ✓

Neutrinos/exotic particle predicted by supersymmetry. ✓

9 a The Jeans criterion states that cloud of gas will begin to collapse under its own gravitation. ✓

When the gravitational potential energy of the cloud exceeds the total random kinetic energy of its particles. ✓

b Blue line as shown. ✓



c i Use $M = Nm$ to eliminate N so that $\frac{GM^2}{R} \approx \frac{3M}{2m}kT$ or $\frac{GM}{R} \approx \frac{3kT}{2m}$ ✓

Now eliminate the mass M through: $M = \rho V = \rho \frac{4\pi R^3}{3}$ so that $\frac{G\rho 4\pi R^3}{3R} \approx \frac{3kT}{2m}$ ✓

Cancelling powers of R and simplifying gives the result $\left(R^2 \approx \frac{9kT}{8\pi mG\rho}\right)$ ✓

ii $R^2 \approx \frac{9kT}{8\pi G\rho m} = \frac{9 \times 1.38 \times 10^{-23} \times 100}{8\pi \times 6.67 \times 10^{-11} \times 1.8 \times 10^{-19} \times 2.0 \times 10^{-27}}$ ✓

$R \approx 1.4 \times 10^{17} \text{ m}$ ✓

10 a i $v = H_0 R$ ✓

ii $E = \frac{1}{2}mv^2 - \frac{GMm}{R}$ ✓

$E = \frac{1}{2}mH_0^2 R^2 - \frac{G\rho 4\pi R^3 m}{3R} = \frac{1}{2}mH_0^2 R^2 - \frac{G\rho 4\pi R^2 m}{3}$ ✓

Factoring gives the result.

iii To escape to infinity the total energy must be zero. ✓

This means that $H_0^2 - \frac{G\rho 4\pi}{3} = 0$. ✓

From which the result follows.

iv $\rho = \frac{3 \times \left(\frac{68 \times 10^3}{10^6 \times 3.09 \times 10^{16}}\right)^2}{8\pi \times 6.67 \times 10^{-11}}$ ✓

$\rho \approx 9 \times 10^{-27} \text{ kg m}^{-3}$ ✓

b In cosmological models with matter density parameters ρ_m and dark energy density ρ_Λ the significance of the critical density is that when $\rho_m + \rho_\Lambda = \rho_c$. ✓

The geometry of the universe is flat, i.e. it has zero curvature. ✓

11 a The CMB is very isotropic which means that we observe the same spectrum in every direction. ✓

However there are small deviations from perfect isotropy in the sense that, in different directions, the temperature deviates from the mean temperature of $T = 2.723 \text{ K}$ by very small amounts (of order

$\frac{\Delta T}{T} \approx 10^{-5}$). ✓

b These deviations are significant because fluctuations in temperature imply fluctuations in density. ✓

And these are required if structures are to develop in the universe. ✓

They are also significant because the magnitude of the fluctuations depends on the geometry of the universe. ✓

Hence study of the fluctuations places limits on the geometry of the universe. ✓

12 a Elements are produced by nuclear fusion. ✓

And nuclear fusion becomes energetically impossible past the peak of the binding energy per nucleon curve which is at iron. ✓

b These are produced mainly by neutron capture. ✓

Nuclei may absorb neutrons which are abundant in a supernova. ✓

As these decay by beta decay. ✓

Nuclei with higher atomic number than iron are produced. ✓

c The CNO cycle requires the fusion of nuclei of carbon, nitrogen and oxygen and since these have a high atomic number the Coulomb barrier that must be overcome is larger than that for hydrogen and helium. ✓

And this requires higher temperatures that are found in the more massive stars. ✓

Answers to test yourself questions

Option A

A1 The beginnings of relativity

- 1 No, some things may be “relative” in relativity but not all is relative. For one thing, a rotating earth means that it is not in an inertial frame.
- 2 The acceleration must be very small so as to be negligible. The acceleration due to the rotation on its axis is $a = 0.034 \text{ m s}^{-2}$ and that due to the rotation around the sun is $a = 5.95 \times 10^{-3} \text{ m s}^{-2}$. In addition, to really consider an observer on earth as being in a true inertial frame there can be no gravity and this can happen only in a frame of reference that is freely falling above the surface of the earth. In such a frame, there is no gravity.
- 3 You can think of many such experiments. One is to let a ball drop from rest. The ball will fall vertically down (as far as you are concerned) in exactly the same way as if the train were at rest.
- 4 No we cannot, since the galvanometer would show the same current irrespective of whether it is the coil or the magnet that moves with respect to the ground.
- 5 You can hang a pendulum from the ceiling. If the train accelerates, the string will not be vertical. If the string is displaced in a given direction, the direction of acceleration will be opposite to that direction.
- 6 The surface of water in a bucket in a rotating frame of reference would not be flat.
- 7 If one inertial observer measures that there is a force, and hence acceleration, other inertial observers must agree. According to the observer moving along with the proton, the proton is at rest so it cannot have a magnetic force on it ($F = qvB$ and $v = 0$). So the electric charges in the wire must exert an electric force that is equal to the magnetic force (in magnitude and direction) it experiences according to an observer at rest with respect to the wire.
- 8 **a** $x' = x - vt$
 $= 20 - 15 \times 5.0$
 $= -55 \text{ m}$
Time is absolute in pre-Einstein physics and so $t' = t = 5.0 \text{ s}$.
b $u = u' + v = 5.0 + 15 = 20 \text{ m s}^{-1}$
- 9 **a** $x = x' + vt$
 $= 24 + (-25) \times 5.0$
 $= -101 \text{ m}$
Time is absolute in pre-Einstein physics and so $t' = t = 5.0 \text{ s}$.
b $u' = u - v = -15 - (-25) = 10 \text{ m s}^{-1}$

A2 The Lorentz transformations

- 10 **a** The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.90^2}} = 2.294$. The time interval of 5.0 min is a proper time interval for the earth observer and so for the Zenga invader the time interval is $\gamma \times 5.0 = 2.294 \times 5.0 = 11.47 \approx 11 \text{ min}$.
b 11 minutes, by exactly the same argument as in **a**.
- 11 The length of the cube in the direction of motion is contracted and so the volume of the cube decreases. The density therefore increases. The density will be $\gamma\rho$.
- 12 The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.95^2}} = 3.20$. The train observers measure the proper time interval. So the ground observers measure $\gamma \times 1.0 = 3.20 \times 1.0 = 3.2 \text{ s}$.

13 The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.95^2}} = 3.20$. The 100 m is the length contracted distance and so the length at rest (the proper length) is $100\gamma = 3.20 \times 100 = 320$ m.

14 a The gamma factor is $\gamma = \frac{30}{28} = 1.07$. The speed is then

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \Rightarrow 1 - \frac{v^2}{c^2} = \frac{1}{\gamma^2} \Rightarrow \frac{v}{c} = \sqrt{1 - \frac{1}{\gamma^2}} = \sqrt{1 - \frac{1}{1.07^2}} = 0.36.$$

- b** Since the trains are identical the proper length of train A is 30 m, which B will measure to be length contracted to 28 m.
c Obviously, 30 m.

15 a The interval of 5.0×10^{-8} s is a proper time interval. The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.95^2}} = 3.20$ so the time interval for the observer in the lab is $\gamma \times 5.0 \times 10^{-8} = 3.20 \times 5.0 \times 10^{-8} = 1.6 \times 10^{-7}$ s.

b The distance traveled is $vt = 0.95 \times 3.0 \times 10^8 \times 1.6 \times 10^{-7} = 45.6 \approx 46$ m.

16 a The time is $\frac{x}{v} = \frac{50 \text{ ly}}{0.995c} \approx 50.3$ yr.

b The time taken according to the spacecraft clocks will be the proper time and this is $\frac{50.3}{\gamma}$ yr. The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.995^2}} \approx 10$. Hence the time is $\frac{50.3}{10} = 5.03$ yr. The students are just over 23 years old when they get to Vega.

17 a The time interval of 4.0 years is the proper time interval for the rocket observers for the events: spacecraft leaves earth and spacecraft sends signal. The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.60^2}} = 1.25$. Hence according to the

earth clocks the signal is sent after $\gamma \times 4.0 = 1.25 \times 4.0 = 5.0$ yr. During this time the spacecraft has traveled a distance $x = vt = 0.60c \times 5.0 = 3.0$ ly according to earth. This distance will be covered at the speed of light by the signal and so it will arrive on earth after 3.0 years according to earth.

b For the rocket, the earth is at a distance of $x' = vt' = 0.60c \times 4.0 = 2.40$ ly. In the time T it takes the signal to get to earth the earth moved away a distance of $x = vt = 0.60c \times T$ so $cT = 0.60c \times T + 2.4 \Rightarrow T = \frac{2.4}{0.4} = 6.0$ yr.

18 The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.80^2}} = \frac{5}{3}$.

a The distance of 8.0 ly is covered at the speed of light and so takes 8.0 years according to earth.

b The distance separating the spacecraft from the space station is $\frac{8.0}{5/3} = 4.8$ ly according to the spacecraft.

Therefore the space station covers this distance in $\frac{4.8 \text{ ly}}{0.80c} = 6.0$ yr.

c The spacecraft is 8.0 ly away from earth (according to earth) when the signal is emitted. In the 8.0 years it takes the signal to arrive at earth the spacecraft moved an additional distance of $0.80c \times 8.0 = 6.4$ ly. The reply signal will cover a distance cT where T is the required arrival time. Then (since spacecraft will travel a distance of $0.80cT$ in the meantime) $cT = 8.0 + 6.4 + 0.80cT \Rightarrow T = 72$ yr.

d The time from the emission of the signal by the spacecraft and its reception is a proper time interval for the spacecraft. According to earth the time interval between these two events is $8.0 + 72 = 80$ yr. Hence the time for the spacecraft is $\frac{80}{5/3} = 48$ yr.

19 a According to the ground the light signal will take time T . In this time the rocket will move a distance νT closer to the mirror. Hence, $cT = D + (D - \nu T) \Rightarrow T = \frac{2D}{c + \nu} = \frac{4.8 \times 10^{12}}{1.90 \times 3.0 \times 10^8} = 8.42 \times 10^3 \approx 8.4 \times 10^3$ s.

b The time for the rocket is the proper time interval since the signal is emitted and received at the same place.

The gamma factor is $\gamma = \frac{1}{\sqrt{1 - \frac{\nu^2}{c^2}}} = \frac{1}{\sqrt{1 - 0.90^2}} = 2.29$. Hence $T' = \frac{8.42 \times 10^3}{2.29} = 3.7 \times 10^3$ s.

20 a $u' = \frac{u - \nu}{1 - \frac{u\nu}{c^2}}$, $\nu = 0.6c$, $u = 0.8c$. Then $u' = \frac{0.2c}{1 - 0.8 \times 0.6} = 0.385c$.

b The answer is obviously $-0.385c$ but we can verify this from: $u' = \frac{u - \nu}{1 - \frac{u\nu}{c^2}}$ where now $\nu = 0.8c$, $u = 0.6c$ so that

$$u' = \frac{-0.2c}{1 - 0.8 \times 0.6} = -0.385c.$$

21 a $u' = \frac{u - \nu}{1 - \frac{u\nu}{c^2}}$, $\nu = -0.6c$, $u = 0.8c$. Then $u' = \frac{-1.40c}{1 - 0.8 \times (-0.6)} = -0.946c$.

b The answer is obviously $0.946c$ but we can verify this from: $u' = \frac{u - \nu}{1 - \frac{u\nu}{c^2}}$ where now $\nu = 0.8c$, $u = -0.6c$ so that

$$u' = \frac{1.40c}{1 - (-0.6) \times 0.8} = 0.946c.$$

22 Here we need to use $u = \frac{u' + \nu}{1 + \frac{u'\nu}{c^2}}$ with $\nu = 0.60c$ and $u' = 0.70c$. This gives $u = \frac{0.70c + 0.60c}{1 + 0.70 \times 0.60} = 0.915c$.

23 Here we need to use $u = \frac{u' + \nu}{1 + \frac{u'\nu}{c^2}}$ with $\nu = -0.60c$ and $u' = 0.70c$. This gives $u = \frac{0.70c + (-0.60c)}{1 + 0.70 \times (-0.60)} = 0.172c$.

24 a The lifetime is $t = \frac{x}{\nu} = \frac{2.00 \times 10^3}{0.95 \times 3.0 \times 10^8} = 7.02 \times 10^{-6}$ s.

b This observers measures the proper time interval between the events muon created and muon decays and so

$$\tau = \frac{t}{\gamma} = \frac{7.02 \times 10^{-6}}{\frac{1}{\sqrt{1 - 0.95^2}}} = 2.19 \times 10^{-6} \text{ s}.$$

25 The lifetime of the pion according to the lab is $t = \frac{20}{\nu}$ and also $t = \tau\gamma = \frac{1}{\sqrt{1 - \frac{\nu^2}{c^2}}} \times 2.6 \times 10^{-8}$.

Hence, $\frac{1}{\sqrt{1 - \frac{\nu^2}{c^2}}} \times 2.6 \times 10^{-8} = \frac{20}{\nu}$. This is best solved on the SOLVER of your GDC. Otherwise,

$\nu^2 \times (2.6 \times 10^{-8})^2 = 20^2 \left(1 - \frac{\nu^2}{c^2}\right)$. This means $\frac{\nu^2}{c^2} \times 60.84 = 400 \left(1 - \frac{\nu^2}{c^2}\right) \Rightarrow \frac{\nu^2}{c^2} = \frac{400}{460.84} \Rightarrow \frac{\nu}{c} = 0.931$ so that

finally $\nu = 2.8 \times 10^8 \text{ m s}^{-1}$.

Note: In the following questions the frames S and S' have their usual meaning i.e. S' moves past S with velocity ν and when the origins coincide clocks are set to zero.

26 $x' = \gamma(x - vt)$ and $t' = \gamma\left(t - \frac{vx}{c^2}\right)$; the gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.75^2}} = 1.5119$. Hence

$$x' = 1.5119 \times (600 - 0.75 \times 3 \times 10^8 \times 2.0 \times 10^{-6}) = 226.8 \approx 230 \text{ m and}$$

$$t' = 1.5119 \times \left(2.0 \times 10^{-6} - \frac{0.75 \times 3 \times 10^8 \times 600}{(3 \times 10^8)^2}\right) = 7.6 \times 10^{-7} \text{ s.}$$

27 $x = \gamma(x' + vt')$ and $t = \gamma\left(t' + \frac{vx'}{c^2}\right)$; the gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.98^2}} = 5.0252$. Hence

$$x = 5.0252 \times (0 + 0.98 \times 3 \times 10^8 \times 6.0 \times 10^{-6}) = 8.9 \times 10^3 \text{ m and}$$

$$t = 5.0252 \times (6.0 \times 10^{-6} - 0) = 3.0 \times 10^{-5} \text{ s.}$$

28 The gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.60^2}} = 1.25$. The S clocks read $t = \frac{120}{0.60 \times 3 \times 10^8} = 6.667 \times 10^{-7} \text{ s}$

when S' passes the 120 m mark in S. Hence with $t' = \gamma\left(t - \frac{vx}{c^2}\right)$ we find

$$t' = 1.25 \times \left(6.667 \times 10^{-7} - \frac{0.60 \times 3 \times 10^8 \times 120}{(3 \times 10^8)^2}\right) = 5.3 \times 10^{-7} \text{ s.}$$

29 a The gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.60^2}} = 1.25$.

i $\Delta x' = \gamma(\Delta x - v\Delta t) = 1.25 \times (1200 - 0.600 \times 3 \times 10^8 \times 6.00 \times 10^{-6}) = 150 \text{ m}$

$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) = 1.25 \times \left(6.00 \times 10^{-6} - \frac{0.600 \times 3 \times 10^8 \times 1200}{(3 \times 10^8)^2}\right) = 4.5 \times 10^{-6} \text{ s}$$

ii The two events are separated by a distance of 1200 m and a time interval of 6.0 μs . A signal from event 1 to event 2 would take $\frac{1200}{6.0 \times 10^{-6}} = 2.0 \times 10^8 \text{ m s}^{-1}$ which is less than the speed of light so event 1 could cause event 2.

b $\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) = 0$ implies $6.0 \times 10^{-6} - \frac{1200v}{c^2} = 0 \Rightarrow v = \frac{6.0 \times 10^{-6} \times 3 \times 10^8}{1200} c = 1.5c$ which is impossible.

30 a The gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.60^2}} = 1.25$.

i $\Delta x' = \gamma(\Delta x - v\Delta t) = 1.25 \times (1200 - 0.600 \times 3 \times 10^8 \times 3.00 \times 10^{-6}) = 825 \text{ m}$

$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) = 1.25 \times \left(3.00 \times 10^{-6} - \frac{0.600 \times 3 \times 10^8 \times 1200}{(3 \times 10^8)^2}\right) = 7.5 \times 10^{-7} \text{ s}$$

ii The two events are separated by a distance of 1200 m and a time interval of 3.0 μs . A signal from event 1 to event 2 would take $\frac{1200}{3.0 \times 10^{-6}} = 4.0 \times 10^8 \text{ m s}^{-1}$ which is more than the speed of light so event 1 could not cause event 2.

b $\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) < 0$ implies $3.0 \times 10^{-6} - \frac{1200v}{c^2} < 0 \Rightarrow v > \frac{3.0 \times 10^{-6} \times 3 \times 10^8}{1200} c > 0.75c$. This says that it is possible to find a frame in which event 2 occurs **before** event 1. In the frame S event 1 occurs before event 2. But this is not a problem since event 1 is not the cause of event 2.

31 The gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.60^2}} = 1.25$,

$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right) = 1.25 \times \left(0 - \frac{0.600 \times 3 \times 10^8 \times 1200}{(3 \times 10^8)^2}\right) = -3.0 \times 10^{-6} \text{ s and so event 1 occurs first}$$

A3 Spacetime diagrams

Note: In the questions that follow the space-time diagrams represent two inertial frames. The black axes represent frame S. The red axes represent a frame S' that moves past frame S with velocity v .

32 Using the dashed line and $v = \frac{x}{t} = \frac{x}{ct}c = \frac{0.6}{1.0}c = 0.6c$.

33 Using the dashed line and $v = \frac{x}{t} = \frac{x}{ct}c = \frac{-0.8}{1.0}c = -0.8c$.

34 a i Blue line

ii Green line

b We must draw the worldline of a photon starting at $x = 1.0$ m and $t = 0$ and see where the worldline intersects the two time axes.

i $ct = 0.61 \text{ m} \Rightarrow t = 2.0 \times 10^{-9} \text{ s}$ in S

ii $ct' \approx 0.50 \text{ m} \Rightarrow t' = 1.7 \times 10^{-9} \text{ s}$ in S'

35 a Using the dashed line and $v = \frac{x}{t} = \frac{x}{ct}c = \frac{1.0}{2.0}c = 0.50c$.

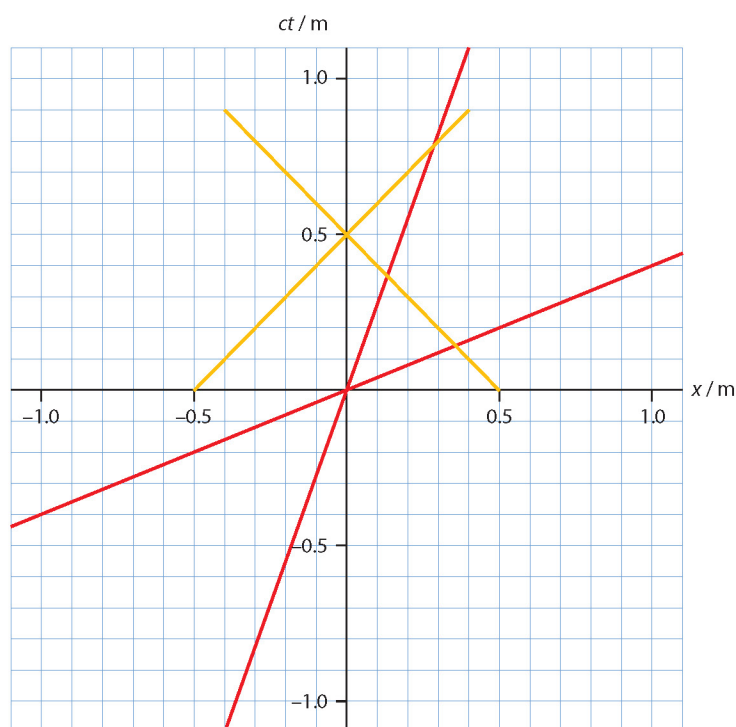
b From the diagram

i $ct = 2.0 \text{ ly} \Rightarrow t = 2.0 \text{ y}$ in S

ii $ct' \approx 1.8 \text{ ly} \Rightarrow t' = 1.8 \text{ y}$ in S'

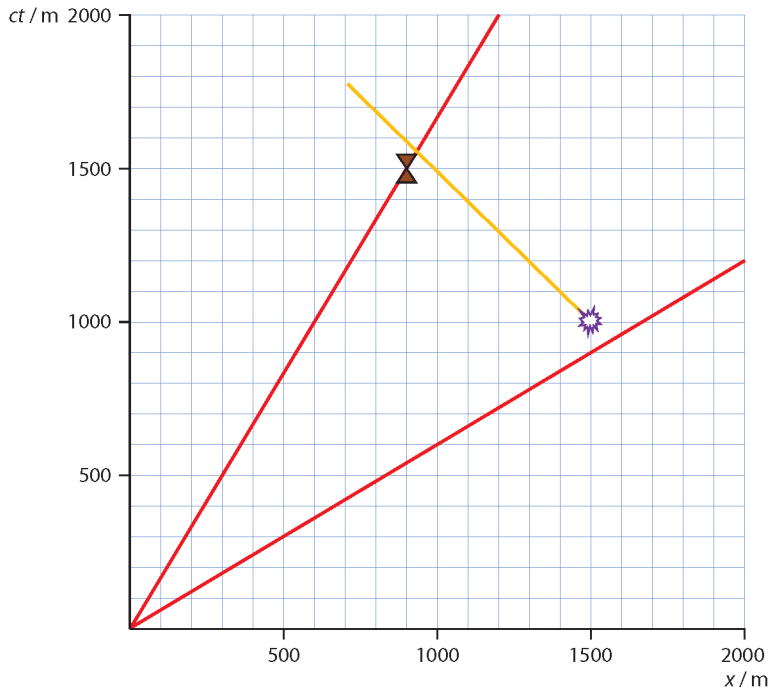
36 a We draw lines from the events parallel to the primed space axis to see that the lamp at $x = +0.5$ m turns on first.

b The lines are 45 lines as shown on the diagram.



c Light from the lamp at $x = +0.5$ m reaches the observer in S' first.

37 a The photon worldline shows that photons arrive after the shield is put on and so the spacecraft is safe.



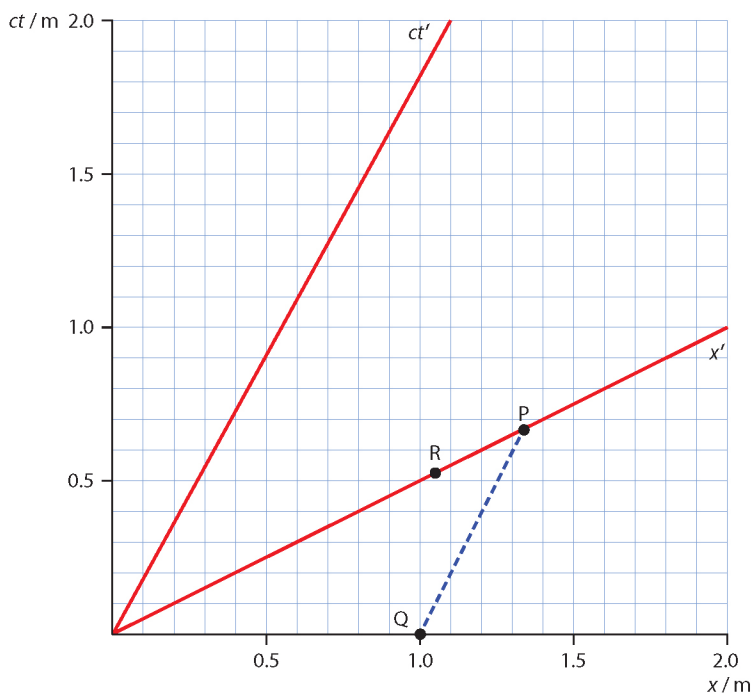
b In the S frame we know that $x = 1500$ m, $ct = 1000$ m. The speed of the spacecraft is $0.60c$ and so the gamma factor is 1.25. Therefore:

i $ct' = c\gamma\left(t - \frac{vx}{c^2}\right) = \gamma\left(ct - \frac{vx}{c}\right) = 1.25 \times \left(1000 - \frac{0.60c \times 1500}{c}\right) = 125$ m. Hence $t' = \frac{125}{3 \times 10^8} = 0.42 \mu\text{s}$

ii $x' = \gamma(x - vt) = 1.25 \times (1500 - 0.60 \times 1000) = 1125$ m

38 a The speed of the primed frame is $0.50c$. The gamma factor is then $\gamma = \frac{1}{\sqrt{1 - 0.50^2}} = 1.1547 \approx 1.2$.

Event P has the same x' coordinate as event Q. The coordinates of Q in S are $x = 1$, $ct = 0$. Hence $x' = \gamma(x - vt) = 1.2 \times (1 - 0) = 1.2$ m. The time coordinate of P is clearly zero. Knowing that the space coordinate of P is about 1.2 m we can measure along the primed x axis to estimate the position of the point with $x' = 1$ m is approximately at point R.



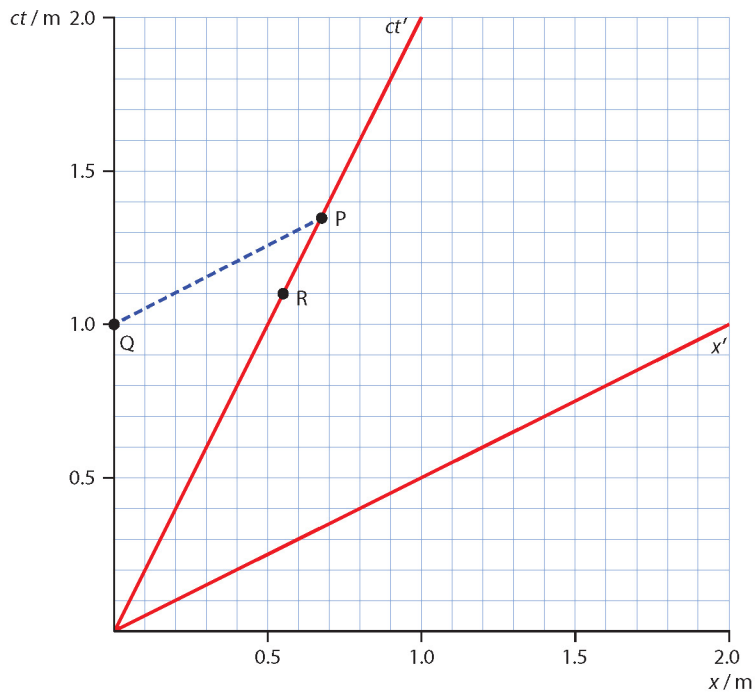
b For the general case again consider event Q that has coordinates in S of $x = 1$ m and $t = 0$. Then $x' = \gamma(x - vt) = \gamma(1 - 0) = \gamma$. $(x', ct') = (\gamma, 0)$

39 a The speed of the primed frame is $0.50c$. The gamma factor is then $\gamma = \frac{1}{\sqrt{1 - 0.50^2}} = 1.1547 \approx 1.2$.

Event P has the same t' coordinate as event Q. The coordinates of Q in S are $x = 0, ct = 1$ m. Hence

$$ct' = c\gamma\left(t - \frac{vx}{c^2}\right) = 1.2 \times (1 - 0) = 1.2 \text{ m.}$$

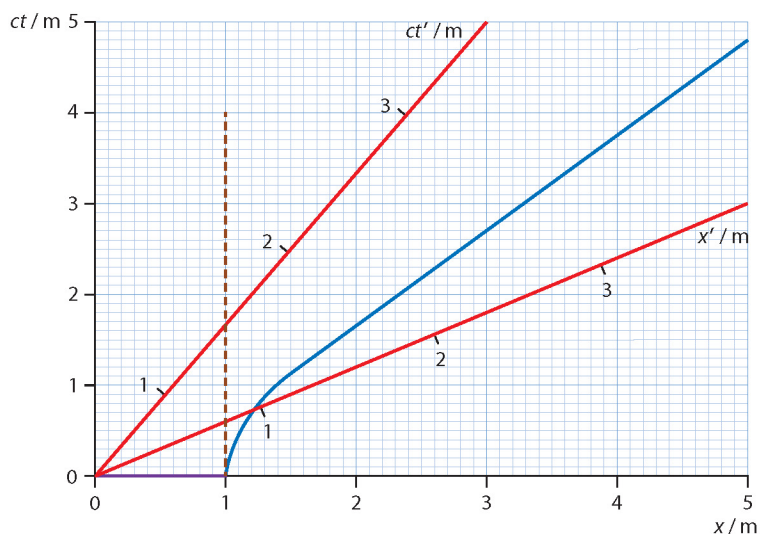
The space coordinate of P is clearly zero. Knowing that the time coordinate of P is about 1.2 m we can measure along the primed t axis to estimate the position of the point with $ct' = 1$ m is approximately at point R.



b For the general case again consider event Q that has coordinates in S of $x = 0$ and $ct = 1$ m. Then

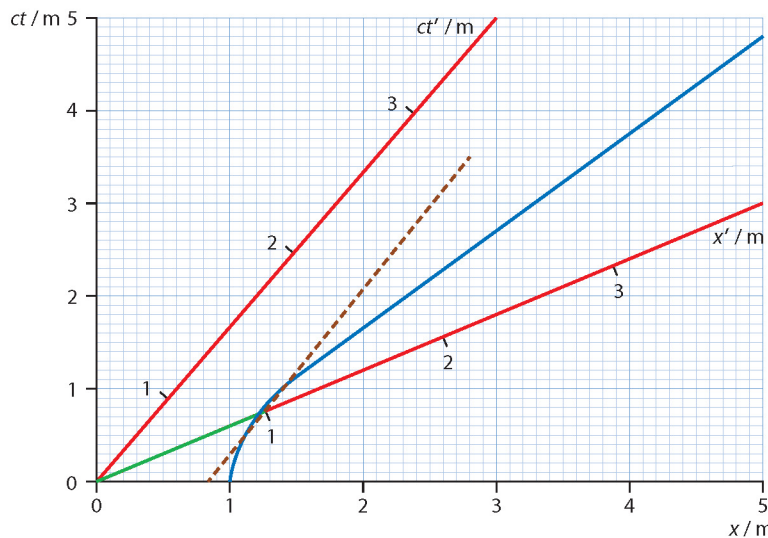
$$ct' = c\gamma\left(t - \frac{vx}{c^2}\right) = \gamma\left(ct - \frac{vx}{c}\right) = \gamma(1 - 0) = \gamma. \text{ P has coordinates } (x', ct') = (0, \gamma).$$

40 a The dotted line is the worldline of the right end of the rod. It intersects the primed x axis at a point that is less than 1 m.



b i See green line segment

ii Dotted line intersects the x axis at a point that is less than 1 m.



41 Draw the dotted line which intersects the time axis at $ct = 1.25$ m. Hence $t = 4.2$ ns.

A4 Relativistic mechanics (HL)

42 a The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.50^2}} = 1.1547$ and so the kinetic energy (and the energy that needs to be supplied) is $E_K = (\gamma - 1)mc^2 = (1.1547 - 1) \times 0.511 = 0.079$ MeV.

b The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.90^2}} = 2.2942$ and so the kinetic energy (and the energy that needs to be supplied) is $E_K = (\gamma - 1)mc^2 = (2.2942 - 1) \times 0.511 = 0.66$ MeV.

c The gamma factor is $\gamma = \frac{1}{\sqrt{1-0.99^2}} = 7.0888$ and so the kinetic energy (and the energy that needs to be supplied) is $E_K = (\gamma - 1)mc^2 = (7.0888 - 1) \times 0.511 = 3.1$ MeV.

43 $E_K = (\gamma - 1)m_0c^2 = 10m_0c^2 \Rightarrow \gamma = 11$. Hence, $11 = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \Rightarrow 1 - \frac{v^2}{c^2} = \frac{1}{11^2} \Rightarrow \frac{v}{c} = \sqrt{1 - \frac{1}{11^2}} = 0.996$.

44 $E_T = 5mc^2 = 5 \times 938 = 4690$ MeV. From $E^2 = (mc^2)^2 + p^2c^2$, $pc = \sqrt{E^2 - (mc^2)^2} = \sqrt{4690^2 - 938^2} = 4595$ MeV. Hence $p = 4.6 \times 10^3$ MeV c^{-1} .

45 The momentum is huge and so the total energy will be just 350 MeV. Explicitly,

$$E = \sqrt{(mc^2)^2 + p^2c^2} = \sqrt{0.511^2 + 350^2} \approx 350 \text{ MeV.}$$

46 $p = 685 \text{ MeV } c^{-1} = \frac{685 \times 10^6 \times 1.6 \times 10^{-19}}{3.0 \times 10^8} = 3.65 \times 10^{-19} \text{ N s}$

47 The total energy is $E = \sqrt{(mc^2)^2 + p^2c^2} = \sqrt{938^2 + 500^2} = 1063$ MeV.

Hence $E_K = 1063 - 938 = 125$ MeV.

- 48 The gamma factor is $\gamma = \frac{200}{135} = 1.4815$. The speed is then

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$1 - \frac{v^2}{c^2} = \frac{1}{1.4815^2}$$

$$\frac{v^2}{c^2} = 0.54438$$

$$v = 0.738c$$

- 49 The total energy will be $E = \sqrt{(mc^2)^2 + p^2c^2} = \sqrt{(0.938)^2 + 1200^2} \approx 1200$ GeV. The kinetic energy is $E_K = 1200 - 0.938 = 1199.1$ GeV. The accelerating voltage to 2 s.f. is then 1200 GV.

- 50 The gamma factor is $\gamma = \frac{1}{\sqrt{1 - 0.99^2}} = 7.0888$. The momentum is then

$$p = \gamma mv = \frac{\gamma mc^2 v}{c^2} = \frac{7.0888 \times 938 \times 0.99c}{c^2} \text{ MeV}$$

$$= \frac{7.0888 \times 938 \times 0.99}{c} \text{ MeV}$$

$$= 6.58 \times 10^3 \text{ MeV } c^{-1}$$

$$\approx 6.6 \times 10^3 \text{ MeV } c^{-1}$$

A faster alternative is: after getting the gamma factor find the total energy to be

$$E = \gamma mc^2 = 7.0888 \times 938 = 6649 \text{ MeV and then from}$$

$$E^2 = (mc^2)^2 + p^2c^2 \Rightarrow pc = \sqrt{6649^2 - 0.938^2} \approx 6.6 \times 10^3 \text{ MeV.}$$

- 51 The energy is $E = \sqrt{m^2c^4 + p^2c^2} = \sqrt{0.938^2 + 1.5^2} = 1.7691$ GeV. The gamma factor is therefore

$$\gamma = \frac{1.7691}{0.938} = 1.8861 \text{ and the speed is:}$$

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$1 - \frac{v^2}{c^2} = \frac{1}{1.8861^2}$$

$$\frac{v^2}{c^2} = 0.71889$$

$$v = 0.848c$$

- 52 a The work done is

$$W = Fd = eEd = 1.6 \times 10^{-19} \times 5.0 \times 10^6 \times 10^3 = 8.0 \times 10^{-10} \text{ J} = 5.0 \times 10^9 \text{ eV} = 5.0 \times 10^3 \text{ MeV. Therefore}$$

$$E_K = 5.0 \times 10^3 \text{ MeV (or 5.0 GeV).}$$

- b The total energy is $E_T = (938 + 5.0 \times 10^3) = 5.938 \times 10^3$ MeV. Hence $\gamma = \frac{E_T}{m_0c^2} = \frac{5.938 \times 10^3}{938} = 6.33$.

$$\text{Hence the speed is: } 6.33 = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \Rightarrow 1 - \frac{v^2}{c^2} = \frac{1}{6.33^2} \Rightarrow \frac{v}{c} = \sqrt{1 - \frac{1}{6.33^2}} = 0.987.$$

53 a From $p = \gamma mv$ and $E = \gamma mc^2$ we get by dividing side by side to get rid of the gamma factor: $\frac{p}{E} = \frac{v}{c^2}$.

Hence $v = \frac{pc^2}{E} = \frac{pc^2}{\sqrt{m^2c^4 + p^2c^2}}$ as required.

b For the electron with momentum $1.00 \text{ MeV } c^{-1}$, $\sqrt{m^2c^4 + p^2c^2} = \sqrt{0.511^2 + 1.00^2} = 1.123$. For the proton, $\sqrt{m^2c^4 + p^2c^2} = \sqrt{938^2 + 1.00^2} \approx 938$. Hence the ratio of the speeds is $\frac{v_e}{v_p} = \frac{938}{1.123} = 835$.

c For a momentum of $1.00 \text{ GeV } c^{-1}$, for the electron $\sqrt{m^2c^4 + p^2c^2} = \sqrt{0.511^2 + (10^3)^2} \approx 10^3$ and for the proton $\sqrt{m^2c^4 + p^2c^2} = \sqrt{938^2 + (10^3)^2} = 1371$ so that $\frac{v_e}{v_p} = \frac{1371}{1000} = 1.37$.

d As the momentum increases we may neglect the rest energy in which case both speeds tend to become the speed of light and so the ratio approaches 1.

54 a The momentum of the fragments must be zero since the original momentum before the breakup was zero. The gamma factor is $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - 0.85^2}} = 1.8983$ and so the momentum

is $1.8983 \times 125 \text{ MeV } c^{-2} \times 0.85c = 201.7 \text{ MeV } c^{-1}$. This is also the momentum of the other fragment. Hence the total energy of each of the fragments is $E_1 = \sqrt{125^2 + 201.7^2} = 237.3 \text{ MeV}$ and $E_2 = \sqrt{250^2 + 201.7^2} = 321.2 \text{ MeV}$. The heavier fragment has gamma factor given by $E_2 = \gamma mc^2 = \gamma \times 250 \text{ MeV} = 321 \text{ MeV}$ and so $\gamma = 1.284$. Hence the speed is:

$$1.284 = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \Rightarrow 1 - \frac{v^2}{c^2} = \frac{1}{1.284^2} \Rightarrow v = 0.627c.$$

b The total energy of the system therefore is $237.3 + 321.2 = 558 \text{ MeV}$. This is the rest energy of the particle that broke up and so its rest mass is $558 \text{ MeV } c^{-2}$.

55 a The total momentum of the electron – positron pair is zero. If only one photon is produced it will have momentum violating the law of momentum conservation.

b Again because of momentum conservation.

c The total energy of the electron is $E = mc^2 + E_K = 0.51 + 2.0 = 2.51 \text{ MeV}$. The positron has the same energy. The total energy is then $E_T = 2 \times 2.51 \approx 5.0 \text{ MeV}$. The photons must have the same energy because they move in opposite directions with the same momentum (magnitude) and hence the same wavelength. So each has an energy of about 2.5 MeV .

56 The gamma factor is $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - 0.80^2}} = \frac{5}{3}$ and so the total energy

of the pion is $E_T = \frac{5}{3} \times 135 = 225 \text{ MeV}$. The momentum of the pion is:

$225 = \sqrt{(135)^2 + (pc)^2} \Rightarrow pc = \sqrt{225^2 - 135^2} = 180 \text{ MeV} \Rightarrow p = 180 \text{ MeV } c^{-1}$. Conservation of energy and momentum gives:

$$225 = hf_A + hf_B$$

$$180 = \frac{hf_A}{c} - \frac{hf_B}{c}$$

To simplify things set $c = 1$ which is alright since we are going to take a ratio and units will not be important. Then

$$225 = hf_A + hf_B$$

$$180 = hf_A - hf_B$$

Adding, $f_A = \frac{405}{2h}$, subtracting, $f_B = \frac{45}{2h}$. The ratio is therefore $\frac{f_A}{f_B} = \frac{405}{45} = 9.0$.

- 57 The momentum of the two bodies is zero and so the particle they form is produced at rest. The gamma factor is $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - 0.80^2}} = \frac{5}{3}$ and so the momentum of each particle is $\frac{5}{3} \times 3.0 \times 0.80 \times 3 \times 10^8 = 1.2 \times 10^9$ N s.

Hence the total energy of each of the particles is $E = \sqrt{(3.0 \times 9.0 \times 10^{16})^2 + (1.2 \times 10^9 \times 3 \times 10^8)^2} = 4.5 \times 10^{17}$ J.

The rest energy of the particle that is formed and so its rest mass is therefore $2 \times 4.5 \times 10^{17} = 9.0 \times 10^{17}$ J and

hence the rest mass is $\frac{9.0 \times 10^{17}}{9.0 \times 10^{16}} = 10$ kg.

58 **a** $p = \gamma m_0 v$

b $E = \gamma m_0 c^2$

c Eliminating the gamma factor from **a** and **b** we get $\frac{p}{E} = \frac{v}{c^2}$. Hence $v = \frac{pc^2}{E}$ as required.

d For a massless particle, $E = \sqrt{0 + p^2 c^2} = pc$ hence $v = \frac{pc^2}{E} = \frac{pc^2}{pc} = c$.

- 59 **a** The work done on the particle is qV and this goes into increasing the kinetic energy of the particle.

I.e. $qV = (\gamma - 1)mc^2$ from which the result follows: $\gamma - 1 = \frac{qV}{mc^2} \Rightarrow \gamma = 1 + \frac{qV}{mc^2}$.

b The gamma factor is $\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - 0.998^2}} = 15.819$. Hence from (a) $15.819 = 1 + \frac{1 \times V}{938} \Rightarrow V = 13.9$ GV

- 60 Let the electron be at rest and assume that a photon of energy E is absorbed by the electron. The momentum of the photon is $p = \frac{E}{c}$ and this will be the momentum of the electron after absorption. The total energy of the electron after absorption is thus $\sqrt{(mc^2)^2 + p^2 c^2} = \sqrt{(mc^2)^2 + E^2}$. By conservation of energy we must also have that this total energy equals the total energy of the system before absorption which is $mc^2 + E$. Therefore:

$mc^2 + E = \sqrt{(mc^2)^2 + E^2}$. Squaring gives

$$(mc^2)^2 + 2Emc^2 + E^2 = (mc^2)^2 + E^2$$

$$2Emc^2 = 0$$

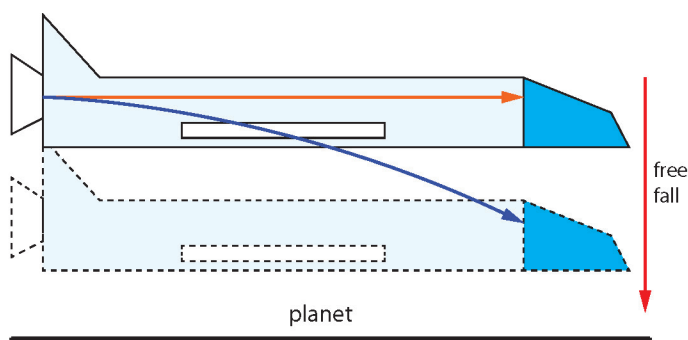
$$E = 0$$

which is impossible: a photon cannot have zero energy. Therefore the assumption that the electron could absorb the photon is false. The case of emitting the photon is similar.

A5 General relativity (HL)

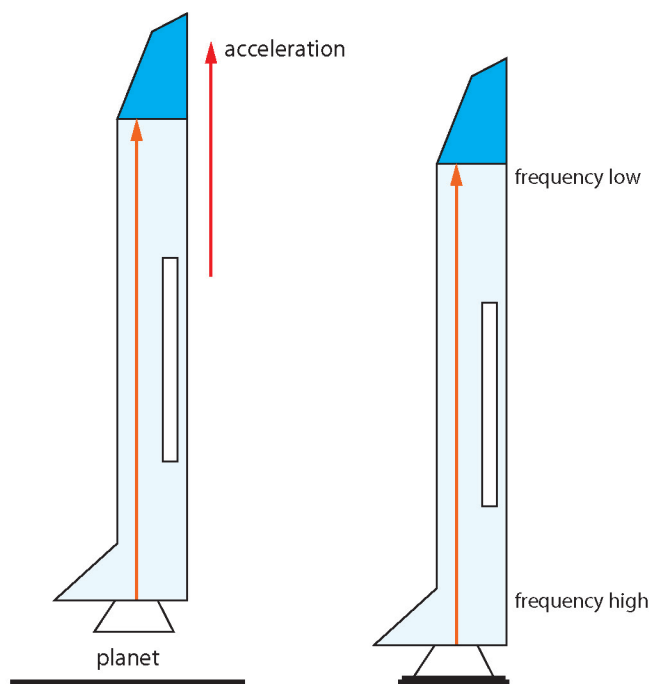
- 61 This is a true statement. Light follows the geodesics of spacetime. In the absence of matter, the spacetime is flat and the geodesics are the ordinary straight lines. In a curved spacetime the "straight lines" i.e. the geodesics look bent because we are prejudiced into thinking in terms of flat spacetimes.

- 62 By the equivalence principle, this frame of reference is equivalent to one which is at rest in a gravitational field. The gravitational field strength is then directed to the left. The helium balloon will “rise” i.e. move opposite to the gravitational field. I.e. it will move to the right.
- 63 For exactly the same reasons as in problem 62 the flame will bend to the right.
- 64 a The equivalence principle states that inertial effects (i.e. effects due to acceleration) cannot be distinguished from effects of gravitation. More precisely, it states that an accelerating frame of reference in outer space is equivalent to a frame of reference at rest in a uniform gravitational field whose field strength is the same as the acceleration of the other frame. It also states that a freely falling frame of reference in a gravitational field is equivalent to an inertial frame of reference.
- b i Consider a rocket that is **freely falling** in a gravitational field. According to observers inside the rocket, the ray of light that is emitted from the back wall of the rocket will travel on a straight line and hit the front of the rocket at a point that has the same distance from the floor as the point of emission (path of light shown in the orange line). This is because to the occupants of the rocket, the rocket is equivalent to a truly inertial frame of reference.



But the rocket is seen to be falling by an observer outside. By the time the light ray goes across, the rocket has fallen and so the ray appears to be following the curved path shown in blue. Thus the outside observer claims that **in a gravitational field, light bends** towards the mass causing the field.

- ii The diagram on the left shows a rocket accelerating in outer space. The diagram to the right shows a rocket at rest on a massive body which is thus equivalent to the first frame. A ray of light is emitted from the back of the rocket and is received at the front.



To an observer outside the rocket on the left, the front of the rocket is moving away from the light ray and so there should be a Doppler redshift. The observer outside expects that the frequency of light measured at the reception point should be smaller than that at emission. Hence the outside observer must conclude that **as the ray of light moves higher in the gravitational field it suffers a redshift**. But frequency is the number of wavefronts received per second so how can the frequency change? The answer has to be that when one second goes by, according to a clock at the base, more than a second goes by, according to a clock at the top, i.e. the equivalence principle predicts **gravitational time dilation**: the interval of time between two events is longer when measured by a clock far from the gravitational field compared to a clock near the gravitational field.

- 65 As the radius gets smaller, a time will be reached when the radius of the object becomes equal to the Schwarzschild radius. The bending of space around the object will be substantial and the object will become a black hole.
- 66 The period of oscillation of the mass at the end of a spring is $T = 2\pi\sqrt{\frac{m}{k}}$. So the acceleration of gravity, i.e. the gravitational field strength does not enter. Hence the period will be the same in **a** and **b**.
- 67 The emitted frequency is $f = \frac{3.00 \times 10^8}{500.0 \times 10^{-9}} = 6.00 \times 10^{14}$ Hz. The shift is found from
- $$\frac{\Delta f}{f} = \frac{gH}{c^2} \Rightarrow \Delta f = 6.00 \times 10^{14} \times \frac{9.81 \times 50.0}{(3.00 \times 10^8)^2} = 3.27 \text{ Hz.}$$
- 68 **a** As the signal moves away from the gravitational field of the star the frequency decreases by the gravitational redshift effect. Hence the wavelength increases.
- b** Time slows down near a collapsed star and so the time in between reception of the signals by the distant spacecraft increases so the frequency of reception decreases.
- c** For the same reason the duration of the pulses increases.
- 69 The acceleration experienced by the clock on the circumference will be greater. By the equivalence principle this clock will behave as an identical clock in a gravitational field. It will therefore run slow relative to a clock in a smaller gravitational field.
- 70 The mass of the earth is $M = 6.0 \times 10^{24}$ kg. The Schwarzschild radius of the earth is
- $$R = \frac{2GM}{c^2} = \frac{2 \times 6.67 \times 10^{-11} \times 6.0 \times 10^{24}}{9.0 \times 10^{16}} = 8.89 \times 10^{-3} \text{ m. Its density would then be}$$
- $$\rho = \frac{M}{\frac{4\pi R^3}{3}} = \frac{6.0 \times 10^{24}}{\frac{4\pi(8.89 \times 10^{-3})^3}{3}} = 2.0 \times 10^{30} \text{ kg m}^{-3}. \text{ (This is larger than nuclear densities by a factor of } 10^{13}.)$$
- 71 From $R = \frac{2GM}{c^2}$, $R = \frac{2 \times 6.67 \times 10^{-11} \times 2.0 \times 10^{31}}{(3.0 \times 10^8)^2} = 3.0 \times 10^4 \text{ m.}$
- 72 A geodesic is a curve in spacetime which has the least length compared to any other curve with the same beginning and end points. A particle upon which the net force is zero follows a geodesic.
- 73 **a** In Newtonian mechanics the path would be explained by saying that a gravitational attractive force acts on the particle changing the original path in a curved path around the massive object.
- b** In relativity the path is explained by saying that the particle follows the geodesic in the curved spacetime around the massive object.

- 74 **a** From the point of view of an observer inside the spacecraft the ball will move on a straight line parallel to the floor and therefore hit the opposite wall at the same height as that at the point of launch. (From the point of view of an inertial observer outside with respect to whom the spacecraft moves upwards, the path will also be a straight line with an upward slope.)
- b** The situation is equivalent to a frame of reference in a gravitational field. Therefore the ball will curve towards the floor following a parabolic path and will hit the opposite wall lower.
- 75 **a** From the point of view of an inertial observer outside the spacecraft the light ray will travel along the original straight line as before entering the spacecraft hitting the opposite wall of the spacecraft at a point closer to the floor than at the point of entry. From the point of view of an observer inside the spacecraft the ray will also move along a straight line with a downward slope.
- b** The frame is now equivalent to a frame of reference at rest on the surface of a massive body. The ray of light will follow a curved path hitting the opposite wall of the spacecraft at a point closer to the floor than at the point of entry and lower than the answer in **a**.
- 76 **a** Because rays of light coming from low in the sky will be bent, these will never reach the observer. The observer sees that his horizon is rising and he can only see things within a vertical cone whose angle is decreasing.
- b** Once inside the event horizon the observer can only see rays of light falling “vertically” into the black hole, i.e. only along a single line.
- 77 The plane is flying at essentially a constant height and so on the surface of a sphere. This is curved and so the plane follows the geodesics of the sphere (these are great circles – circles whose plane goes through the center of the earth) because these have the least length so the least amount of fuel is being used.
- 78 Einstein thought of the contraption in free fall. Being in free fall, the brass ball is effectively weightless since, by the equivalence principle, it is equivalent to a ball in zero gravitational field. The spring will then pull the mass in the bowl. Einstein was very proud of his present and showed it to all who visited him.

79 We have that $\Delta t = \frac{\Delta t_0}{\sqrt{1 - \frac{R_S}{r}}}$, i.e. $2.0 = \frac{1.0}{\sqrt{1 - \frac{R_S}{r}}}$ and so $1 - \frac{R_S}{r} = \frac{1}{4}$, $\frac{R_S}{r} = \frac{3}{4}$ giving finally $r = \frac{4}{3}R_S$. The

observer is at distance of $r = \frac{4}{3}R_S$ from the event horizon.

- 80 It will take longer since clocks near massive objects run slow compared to clocks far away. The accelerating spacecraft is equivalent to one at rest in a gravitational field.

- 81 **a** A black hole is a singularity in spacetime, a point of infinite curvature.

b $R = \frac{2GM}{c^2} = \frac{2 \times 6.67 \times 10^{-11} \times 5.0 \times 10^{35}}{9.0 \times 10^{16}} = 7.4 \times 10^8 \text{ m.}$

- c** This radius is the distance from the black hole where the escape speed is equal to the speed of light. The black hole does not have a radius since it is a point.

d The observer next to the source measures a period of $T = \frac{1}{f} = \frac{1}{7.50 \times 10^{14}} = 1.3 \times 10^{-15} \text{ s.}$

e The distant observer will measure a period of $\Delta t = \frac{\Delta t_0}{\sqrt{1 - \frac{R_S}{r}}} = \frac{1.3 \times 10^{-15}}{\sqrt{1 - \frac{R_S}{1.1R_S}}} = 4.4 \times 10^{-15} \text{ s}$ and hence a

frequency of $f = \frac{1}{4.4 \times 10^{-15}} = 2.3 \times 10^{14} \text{ Hz.}$

82 a $R_S = \frac{2GM}{c^2}$

b The area is $A_S = 4\pi R_S^2 = \frac{16\pi G^2 M^2}{c^4}$.

c Mass constantly falls into the black hole and so the radius keeps increasing. Hence the area also increases.

d Entropy is another quantity that always increases. The black hole has thermodynamic properties (as discovered independently by D. Christodoulou and J. Bekenstein) and in fact behaves as a real thermodynamic black body that radiates according to the Stefan-Boltzmann law. This is because of quantum effects as explained by S. Hawking. The effective “temperature” of the black hole is inversely proportional to its mass. This means that small black holes radiate a lot.

83 a The ray will fall straight into the hole.

b The rays will be bent and enter the observer’s eye.

Answers to test yourself questions

Option B

B1 Rotational dynamics

1 Use $\theta = \frac{(\omega + \omega_0)t}{2}$ to get $\theta = \frac{(15 + 3.5) \times 5.0}{2} = 46.25 \approx 46$ rad.

2 Use $\omega^2 = \omega_0^2 + 2\alpha\theta$ to get $\omega^2 = 5.0^2 + 2 \times 2.5 \times 54$ and so $\omega = 17.18 \approx 17$ rad s⁻¹.

3 Use $\omega^2 = \omega_0^2 + 2\alpha\theta$ to get $12.4^2 = 3.2^2 + 2 \times \alpha \times 20 \times 2\pi$. Hence $\alpha = 0.571 \approx 0.57$ rad s⁻².

4 Let the forces at each support be L (for left) and R (for right). These are vertically upwards. Then

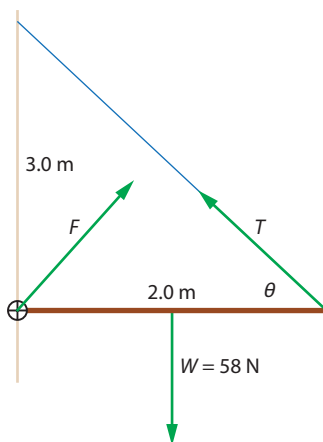
Translational equilibrium: $L + R = 450 + 120 = 570$ N

Rotational equilibrium: $120 \times 2.0 + 450 \times 2.5 = R \times 5.0$ by taking torques about the support at the left.
Then $R = 273 \approx 270$ N and hence $L = 297 \approx 300$ N.

5 Let the forces at each support be L (for left) and R (for right). These are vertically upwards. Then just before the rod tips over we have equilibrium and the left support force tends to become zero since the rod will no longer touch the support. Hence by taking torques about the right support we find:

$30 \times 9.8 \times 0.80 = 40 \times 9.8x$ and $x = 0.60$ m.

6 a The forces are as shown in the diagram below:



Translational equilibrium demands that:

$$T_x = F_x$$

$$T_y + F_y = 58$$

Rotational equilibrium demands that (we take torques about an axis through the point of support at the wall):

$$58 \times 1.0 = T_y \times 2.0 \Rightarrow T_y = 29 \text{ N.}$$

Now $T_y = T \sin \theta$ and from the diagram, $\tan \theta = \frac{3.0}{2.0} = 1.5$. Hence $\theta = 56.31^\circ$. Thus

$$T = \frac{T_y}{\sin \theta} = \frac{29}{\sin 56.31^\circ} = 34.854 \approx 35 \text{ N.}$$

b Since $T_y = 29$ N, $F_y = 58 - 29 = 29$ N also. Finally, $F_x = T_x = 34.854 \cos 56.31^\circ = 19.33$ N. Hence the wall force is $F = \sqrt{F_x^2 + F_y^2} = \sqrt{19.33^2 + 29^2} = 34.854 \approx 35$ N. The angle it makes with the horizontal is

$$\tan^{-1} \frac{29}{19.33} = 56.31^\circ \approx 56^\circ.$$

7 The torque provided by the force about the axis of rotation is $\Gamma = FR$. We know that $\Gamma = I\alpha$.

The moment of inertia of the cylinder is $I = \frac{1}{2}MR^2$ and so $\alpha = \frac{\Gamma}{I} = \frac{2F}{MR} = \frac{2 \times 6.5}{5.0 \times 0.20} = 13 \text{ rad s}^{-2}$.

Therefore $\omega = \omega_0 + \alpha t = 0 + 13 \times 5.0 = 65 \text{ rad s}^{-1}$.

8 Let h be the height from which the bodies are released.

For the point particle will have: $Mgh = \frac{1}{2}Mv^2 \Rightarrow v = \sqrt{2gh}$.

For the others,

$$Mgh = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2$$

$$Mgh = \frac{1}{2}Mv^2 + \frac{1}{2}I \frac{v^2}{R^2}$$

$$2gh = v^2 \left(1 + \frac{I}{MR^2} \right)$$

$$v = \sqrt{\frac{2gh}{1 + \frac{I}{MR^2}}}$$

So

$$\text{Sphere: } v = \sqrt{\frac{2gh}{1 + \frac{2}{5} \frac{MR^2}{MR^2}}} = \sqrt{\frac{2gh}{\frac{7}{5}}} = \sqrt{2gh} \sqrt{\frac{5}{7}}$$

$$\text{Cylinder: } v = \sqrt{\frac{2gh}{1 + \frac{1}{2} \frac{MR^2}{MR^2}}} = \sqrt{\frac{2gh}{\frac{3}{2}}} = \sqrt{2gh} \sqrt{\frac{2}{3}}$$

$$\text{Ring: } v = \sqrt{\frac{2gh}{1 + \frac{MR^2}{MR^2}}} = \sqrt{\frac{2gh}{2}} = \sqrt{2gh} \sqrt{\frac{1}{2}}$$

Hence $v_{\text{ring}} < v_{\text{cylinder}} < v_{\text{sphere}} < v_{\text{point}}$

9 a One revolution corresponds to 2π radians and so the disc is making

$$\omega = \frac{45 \text{ rev}}{2\pi \text{ s}} = \frac{45}{2\pi} \frac{\text{rev}}{\frac{1}{60} \text{ min}} = \frac{45 \times 60}{2\pi} = 429.7 \approx 430 \text{ rpm.}$$

b The angular acceleration has to be $\alpha = \frac{45}{4.0} = 11.25 \text{ rad s}^{-2}$. From

$$\Gamma = I\alpha$$

$$FR = \frac{1}{2}MR^2\alpha$$

we deduce that $F = \frac{1}{2}MR\alpha = \frac{1}{2} \times 12 \times 0.35 \times 11.25 = 23.6 \approx 24 \text{ N}$.

c From $\omega^2 = \omega_0^2 + 2\alpha\theta$ we find $0 = 45^2 + 2 \times (-11.25)\theta$ and so $\theta = 90 \text{ rad}$. This corresponds to $\frac{90}{2\pi} = 14.3 \approx 14$ revolutions.

10 a $\Gamma = I\alpha \Rightarrow MgR = \frac{7}{5}MR^2\alpha$. Thus $\alpha = \frac{5g}{7R}$.

b It is not: the torque of the weight about the axis is decreasing because the perpendicular distance between the weight and the axis is decreasing.

c Conservation of energy: $MgR = \frac{1}{2}I\omega^2$. Hence $MgR = \frac{1}{2} \times \frac{7}{5}MR^2\omega^2$ and so $\omega = \sqrt{\frac{10g}{7R}}$.

11 a $\Gamma = I\alpha \Rightarrow Mg\frac{L}{2} = \frac{1}{3}ML^2\alpha$. Thus $\alpha = \frac{3g}{2L} = \frac{3}{2} \times \frac{9.81}{1.20} = 12.26 \approx 12.3 \text{ rad s}^{-2}$.

b It is not: the torque of the weight about the axis is decreasing because the perpendicular distance between the weight and the axis is decreasing.

c Conservation of energy: $Mg\frac{L}{2} = \frac{1}{2}I\omega^2$. Hence $MgL = \frac{1}{3}ML^2\omega^2$ and $\omega = \sqrt{\frac{3g}{L}} = \sqrt{\frac{3 \times 9.81}{1.20}} = 4.95 \text{ rad s}^{-1}$.

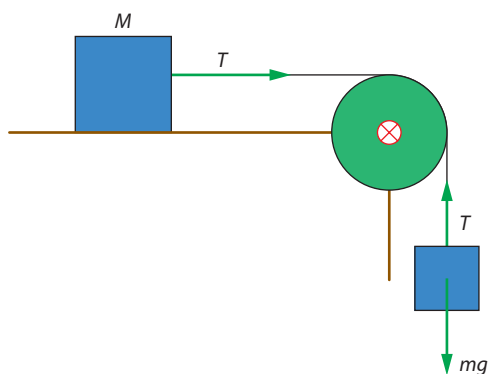
12 From Newton's second law: $F = Ma$ and $Fd = I\alpha = \frac{1}{2}MR^2\alpha$. Combining these two equations gives

$$Mad = \frac{1}{2}MR^2\alpha$$

$$M\alpha Rd = \frac{1}{2}MR^2\alpha$$

$$d = \frac{R}{2}$$

13 a The relevant forces are as shown.



Therefore:

$$mg - T = ma$$

$$T = Ma$$

Hence, adding side by side, $mg = Ma + ma$ and so $a = \frac{mg}{M + m}$.

b Using the hint we now have:

$$mg - T_1 = ma \quad \text{and} \quad (T_1 - T_2)R = I\alpha$$

$$T_2 = Ma$$

Assuming no slipping at the pulley, $\alpha = \frac{a}{R}$ and so $(T_1 - T_2)R = \frac{Ia}{R}$. Adding the first two equations gives

$$mg - (T_1 - T_2) = (m + M)a \quad \text{or} \quad mg - \frac{Ia}{R^2} = (m + M)a \quad \text{i.e.} \quad mg - \frac{1}{2}Ma = (m + M)a \quad \text{so that, finally,} \quad a = \frac{mg}{m + \frac{3}{2}M}$$

14 a The kinetic energy initially is $E_k = \frac{1}{2}I\omega^2 = \frac{1}{2} \times 280 \times \left(\frac{320 \times 2\pi}{60}\right)^2 = 1.57 \times 10^5 \approx 1.6 \times 10^5 \text{ J}$. This is also the work needed to stop the disc.

b The power developed is $P = \frac{E}{t} = \frac{1.57 \times 10^5}{12} = 1.31 \times 10^4 \approx 1.3 \times 10^4 \text{ W}$.

15 X: $E_K = \frac{1}{2}I\omega^2 = \frac{1}{2} \times \frac{1}{12} ML^2 \omega^2 = \frac{1}{24} \times 2.40 \times 1.20^2 \times 4.50^2 = 2.916 \approx 2.92 \text{ J}$;

$$L = I\omega = \frac{1}{12} ML^2 \omega = \frac{1}{12} \times 2.40 \times 1.20^2 \times 4.50 = 1.296 \approx 1.30 \text{ Js.}$$

Y: $E_K = \frac{1}{2}I\omega^2 = \frac{1}{2} \times \frac{1}{3} ML^2 \omega^2 = 1.17 \text{ J}$; $L = I\omega = \frac{1}{3} ML^2 \omega = 5.18 \text{ Js.}$

16 There are no external torques so angular momentum is conserved: $I_1\omega_1 = I_2\omega_2$.

$$\frac{2}{5} MR^2 \omega_1 = \frac{2}{5} M \left(\frac{R}{50} \right)^2 \omega_2 \text{ leading to } \frac{\omega_2}{\omega_1} = 2.5 \times 10^4.$$

17 The angular momentum before and after the ring begins to move must be the same. Before it moves, the angular momentum is zero. After it begins to move it is: $L = I\omega + mvR$.

Hence $I\omega + mvR = 0$, i.e. $\omega = -\frac{mvR}{I} = -\frac{0.18 \times 0.50 \times 0.80}{0.20} = -0.36 \text{ rad s}^{-1}$. (The minus sign indicates that the ring rotates in the opposite direction to that of the car.)

B2 Thermodynamics

18 Use $pV^{\frac{5}{3}} = c$ to get $8.1 \times 10^5 \times (2.5 \times 10^{-3})^{\frac{5}{3}} = p \times (4.6 \times 10^{-3})^{\frac{5}{3}}$ i.e.

$$p = 8.1 \times 10^5 \times \left(\frac{2.5 \times 10^{-3}}{4.6 \times 10^{-3}} \right)^{\frac{5}{3}}$$

$$p = 2.9 \times 10^5 \text{ Pa}$$

19 We know that $pV^{\frac{5}{3}} = c$ and also $pV = nRT$. From the ideal gas law we find $p = \frac{nRT}{V}$ and

substituting in the first equation we find $\frac{nRT}{V} V^{\frac{5}{3}} = c$ or $TV^{\frac{2}{3}} = c'$ (another constant). Hence,

$$560 \times (2.8 \times 10^{-3})^{\frac{2}{3}} = T \times (4.8 \times 10^{-3})^{\frac{2}{3}}. \text{ Hence}$$

$$T = 560 \times \left(\frac{2.8 \times 10^{-3}}{4.8 \times 10^{-3}} \right)^{\frac{2}{3}}$$

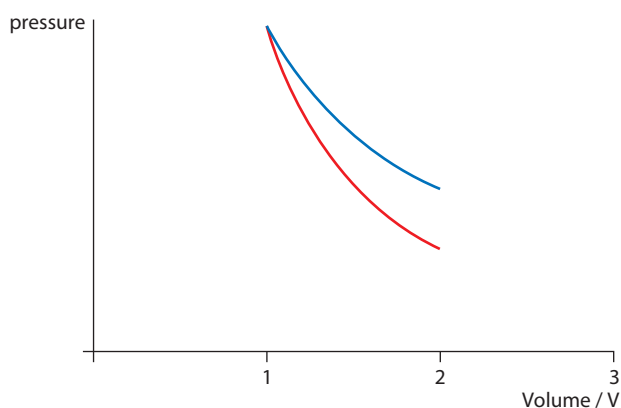
$$T = 391 \approx 390 \text{ K}$$

20 $W = p\Delta V = 4.0 \times 10^5 \times (4.3 - 3.6) \times 10^{-3} = 280 \text{ J.}$

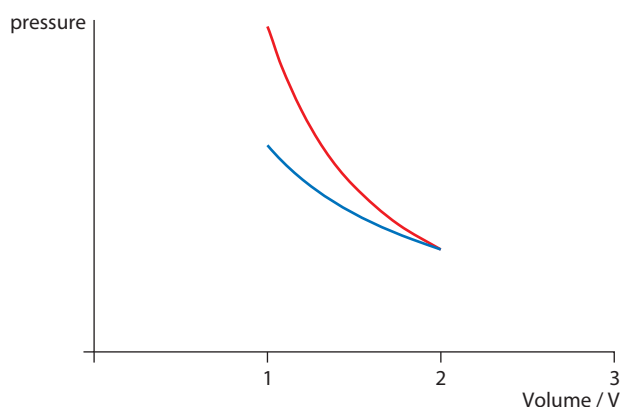
21 a The process is isothermal and so $\Delta U = 0$. Hence $Q = 0 + W = 0 - 6500 = -6500 \text{ J.}$

b The adiabatic compression is steeper than the isothermal and so there is more area under the graph and so more work.

22 Isothermal is the blue curve and the red is adiabatic. The area under the isothermal curve is larger and so the work done is larger.



- 23 Isothermal is the blue curve and the red is adiabatic. The area under the adiabatic curve is larger and so the work done is larger.



	W	ΔU	ΔT
X	positive	zero	zero
Y	zero	positive	positive
Z	positive	positive	positive

- 25 **a** From $Q = \Delta U + W$ and $Q = 0$ we find $\Delta U = -W$. The gas is compressed and so $W < 0$. Hence $\Delta U > 0$. In an ideal gas the internal energy is proportional to temperature (in kelvin) and so increasing U means increasing T .
b The gas is compressed by a piston that is rapidly moved in. The piston collides with molecules and gives kinetic energy to them increasing the average random kinetic energy of the molecules. Hence the temperature increases since temperature is proportional to the average random kinetic energy.

26 **a** Use $W = p\Delta V = 6.00 \times 10^6 \times (0.600 - 0.200) = 2.4 \text{ MJ}$.

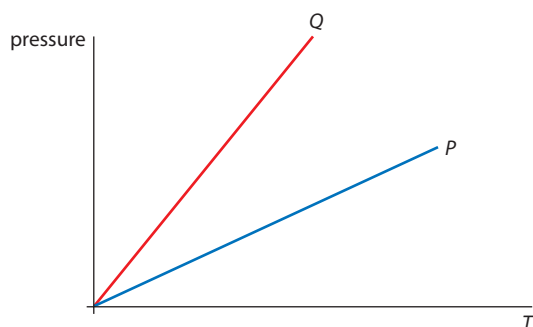
b From the gas law, $\frac{V_1}{T_1} = \frac{V_2}{T_2}$ and so $\frac{0.200}{300} = \frac{0.600}{T_2}$ giving $T_2 = 900 \text{ K}$.

c The change in internal energy is $\Delta U = \frac{3}{2} Rn\Delta T = \frac{3}{2} \frac{pV}{T} \Delta T = \frac{3}{2} \times \frac{6.00 \times 10^6 \times 0.200}{300} = 3.6 \text{ MJ}$.

Therefore $Q = \Delta U + W = 3.6 + 2.4 = 6.0 \text{ MJ}$

- 27 From $Q = \Delta U + W$ we get $\Delta U = Q - W$. So we can have $Q > 0$ but as long as $Q - W < 0$, i.e. when the work done is greater than the heat supplied the temperature will actually drop.

- 28 **a** We must first determine if the gas is expanding or whether it is being compressed. Since $pV = nRT$ it follows that $p = \frac{nRT}{V}$. For constant volume, the graph of pressure versus temperature would be a straight line through the origin. This is not what we have so the volume is changing. The graph below shows two straight lines along each of which the volume is constant. Since $p = \frac{nRT}{V}$, the red line having a bigger slope corresponds to the smaller volume. Hence the volume from P to Q decreases and so $W < 0$.



- b** From $Q = \Delta U + W$ we see that $\Delta U < 0$ since the temperature is decreasing and since the gas is being compressed, we also have that $W < 0$. Hence $Q < 0$ and thermal energy is being removed from the gas.
- 29** For the constant volume case we have $Q = \Delta U + W = \Delta U + 0$. For the constant pressure case we have that $Q = \Delta U' + W$. Since the heats are equal $\Delta U = \Delta U' + W$ which shows that $\Delta U > \Delta U'$ since $W > 0$. The internal energy for an ideal gas is a function only of temperature and so the temperature increase is greater for the constant volume case.
- 30** Working as in the previous problem, $Q_V = \Delta U$ and $Q_p = \Delta U + W$. The two changes in internal energy are the same because the temperature differences are the same. Since $W > 0$, $Q_p > Q_V$.
- 31 a** Steeper curve starting at 1.
- b** Larger area under the adiabatic so more work done.
- c i** $Q = \Delta U + W$ with $\Delta U = 0$ and $W = -25 \text{ kJ}$ so $Q = -25 \text{ kJ}$. Hence the change in entropy for the gas is

$$\Delta S = \frac{Q}{T} = -\frac{25 \times 10^3}{300} = -83.3 \approx -83 \text{ J K}^{-1}$$

- ii** The surroundings receive heat 25 kJ and so their entropy increases by $+\frac{25 \times 10^3}{300} = +83.3 \approx +83 \text{ J K}^{-1}$.
- d** The entropy change for the universe is zero. This is possible only for idealised reversible processes and an isothermal compression is such a process.
- 32 a** An adiabatic curve is a curve on a pressure-volume graph represents a process in which no heat enters or leaves a system.
- b** BC is an isobaric process and DA is isovolumetric.
- c** Along BC the temperature increases and so $\Delta U > 0$. The gas expands and so does work and hence $W > 0$. Therefore, from the first law, $Q = \Delta U + W > 0$ and so heat is given to the gas. Along DA the temperature drops and so $\Delta U < 0$; no work is done and so $Q < 0$. CD and AB are adiabatics and so $Q = 0$. So heat is supplied only along BC.
- d i** Heat is taken out along DA: the temperature at D is

$$T = \frac{pV}{nR} = \frac{4.0 \times 10^6 \times 8.6 \times 10^{-3}}{0.25 \times 8.31} = 1.66 \times 10^4 \approx 1.7 \times 10^4 \text{ K and that at A is}$$

$$T = 1.4 \times \frac{1.66 \times 10^4}{4.0} = 5.80 \times 10^3 \approx 5.8 \times 10^3 \text{ K. Therefore}$$

$$Q = \Delta U + W = \Delta U + 0 = \frac{3}{2} nR\Delta T = \frac{3}{2} \times 0.25 \times 8.31 \times (0.580 - 1.66) \times 10^4 = -3.37 \times 10^4 \approx -3.4 \times 10^4 \text{ J.}$$

- ii** The efficiency is $e = \frac{W}{Q_{\text{in}}} = \frac{Q_{\text{in}} - Q_{\text{out}}}{Q_{\text{in}}}$. Hence $0.36 = \frac{Q_{\text{in}} - 3.37 \times 10^4}{Q_{\text{in}}}$. This gives

$$Q_{\text{in}} = \frac{3.37 \times 10^4}{0.64} = 5.27 \times 10^4 \approx 5.3 \times 10^4 \text{ J.}$$

- iii** The area of the loop is the net work done which equals

$$W = Q_{\text{in}} - Q_{\text{out}} = 5.27 \times 10^4 - 3.37 \times 10^4 = 1.90 \times 10^4 \approx 1.9 \times 10^4 \text{ J.}$$

- 33** Let a quantity of heat Q_1 be supplied to an ideal gas at constant volume: then $Q_1 = nc_V\Delta T$. Similarly, supply a quantity of heat Q_2 to another ideal gas of the same mass at constant pressure so that the change in temperature will be the same. $Q_2 = nc_p\Delta T$. Q_1 and Q_2 will be different temperature because in the second case some of the heat will go into doing work as the gas expands at constant pressure. From the first law of thermodynamics we have that $Q_1 = nc_V\Delta T = \Delta U + 0$ and $Q_2 = nc_p\Delta T = \Delta U + p\Delta V$. Subtracting these two equations gives $nc_p\Delta T - nc_V\Delta T = p\Delta V$. From the ideal gas law we find $p\Delta V = nR\Delta T$ and so $nc_p\Delta T - nc_V\Delta T = nR\Delta T$ or $c_p - c_V = R$ as required.

34 a A and B have the same pressure and so $\frac{V_1}{T_1} = \frac{V_2}{T_2}$ i.e. $\frac{0.1}{800} = \frac{0.4}{T_2}$ hence $T_2 = 3200$ K. B and C have the same volume hence $\frac{P_1}{T_1} = \frac{P_2}{T_2}$ i.e. $\frac{4}{3200} = \frac{2}{T_2}$ hence $T_2 = 1600$ K. C and D have the same pressure and so $\frac{V_1}{T_1} = \frac{V_2}{T_2}$ i.e. $\frac{0.4}{1600} = \frac{0.1}{T_2}$ hence $T_2 = 400$ K.

b A to B: $\Delta U = \frac{3}{2} Rn\Delta T = \frac{3}{2} \frac{pV}{T} \Delta T = \frac{3}{2} \times \frac{4.0 \times 10^5 \times 0.10}{800} \times (3200 - 800) = 180$ kJ

B to C: $\Delta U = \frac{3}{2} Rn\Delta T = \frac{3}{2} \frac{pV}{T} \Delta T = \frac{3}{2} \times \frac{4.0 \times 10^5 \times 0.40}{3200} \times (1600 - 3200) = -120$ kJ

C to D: $\Delta U = \frac{3}{2} Rn\Delta T = \frac{3}{2} \frac{pV}{T} \Delta T = \frac{3}{2} \times \frac{2.0 \times 10^5 \times 0.40}{1600} \times (400 - 1600) = -90$ kJ

D to A: $\Delta U = \frac{3}{2} Rn\Delta T = \frac{3}{2} \frac{pV}{T} \Delta T = \frac{3}{2} \times \frac{2.0 \times 10^5 \times 0.10}{400} \times (800 - 400) = +30$ kJ

c From A to B: $Q = \Delta U + W$. $W = +p\Delta V = 4.0 \times 10^5 \times 0.3 = 120$ kJ. Hence $Q = 180 + 120 = 300$ kJ.

From B to C: $W = 0$ so $Q = -120$ kJ. From C to D: $W = -p\Delta V = 2.0 \times 10^5 \times 0.3 = -60$ kJ. Hence $Q = -90 - 60 = -150$ kJ. Finally from D to A: $\Delta U = 30$ kJ and $W = 0$ so that $Q = +30$ kJ.

d The net work is the area of the loop which is $2.0 \times 10^5 \times 0.3 = 60$ kJ. (As a check this must also be the net heat into the system i.e. $+300 - 120 - 150 + 30 = 60$ kJ.) The heat in is 300 kJ and so the efficiency is $\frac{60}{300} = 0.20$.

35 a The ice is receiving heat; the molecules are moving faster about their equilibrium positions. The disorder increases and so the entropy of the ice molecules increases.

b During melting heat is provided to the ice molecules and so their entropy is again increasing.

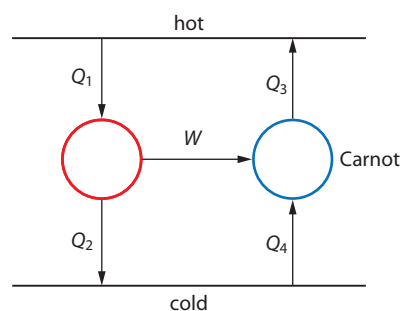
c The water must have had a temperature greater than $+15^\circ\text{C}$. Therefore as the temperature drops to $+15^\circ\text{C}$ the entropy of the water is decreasing. The overall change in entropy of the system must be positive.

36 The maximum possible efficiency for an engine working with these temperatures is, according to Carnot,

$$e = 1 - \frac{300}{500} = 0.40. \text{ So the proposed engine is impossible.}$$

37 a It is impossible for heat to flow from a cold to a warmer place without performing work.

b Imagine the work output of the engine to be the input work in a Carnot refrigerator.



Since the red engine has better efficiency than Carnot we must have that $\frac{W}{Q_1} > \frac{W}{Q_3}$ i.e. that $Q_3 > Q_1$. Notice

that $W = Q_1 - Q_2 = Q_3 - Q_4$ so that $Q_4 - Q_2 = Q_3 - Q_1$ and each of these differences is **positive**. Now the combined effect of the two engines is to remove transfer heat $Q_4 - Q_2$ from the cold reservoir and deposit the (equal amount) $Q_3 - Q_1$ into the hot reservoir without performing work. This violates the Clausius form of the second law. Hence an engine with an efficiency greater than Carnot (for the same temperatures) is impossible.

B3 Fluids (HL)

- 38 Pressure only depends on depth and the pressure exerted on the surface so the pressures here are equal.
- 39 In the second case the fluid exerts an upward force on the floating wood and so the wood exerts an equal and opposite force on the fluid. Hence when put on a scale the second beaker will show a greater reading.
- 40 **a** The second beaker has a smaller mass of water and so a smaller weight by an amount equal to the weight of the displaced water. But there is an additional force pushing down on the liquid that is equal to the buoyant force. This force is equal to the weight of the displaced water and so the readings when the two beakers are put on a scale will be the same.
- b** They points are at the same depth so the pressure is the same.
- 41 Draw a horizontal line through X. Points on this line have the same pressure. Hence

$$P_X = P_{\text{atm}} + \rho gh = 1.0 \times 10^5 + 13600 \times 9.8 \times 0.55 = 1.7 \times 10^5 \text{ Pa}$$
- 42 Since the ice cube floats its weight is equal to the buoyant force. In turn the buoyant force is equal to the weight of the **displaced** water. Hence then the ice cube melts it will have a volume equal to the displaced volume of water and so the level will remain the same.
- 43 From Pascal's principle, in the first case the pressure is $p_0 + \rho gh$. In the second $p_0 + \frac{W}{A} + \rho gh$ and in the third $p_0 + \frac{2W}{A} + \rho gh$.
- 44 The buoyant force on the sphere equals the weight of the sphere and also $B = \rho g V_{\text{imm}}$. Hence

$$W = B = 10^3 \times 9.8 \times \frac{1}{2} \times \frac{4\pi R^3}{3} = 10^3 \times 9.8 \times \frac{1}{2} \times \frac{4\pi \times 0.15^3}{3} = 69.27 \text{ N. The mass is therefore}$$

$$m = \frac{69.27}{9.8} = 7.07 \text{ kg. The volume of the material in the sphere is } \frac{4\pi \times 0.15^3}{3} - \frac{4\pi \times 0.14^3}{3} = 2.64 \times 10^{-3} \text{ m}^3.$$
Hence the density is $\frac{m}{V} = \frac{7.07}{2.64 \times 10^{-3}} = 2.678 \times 10^3 \approx 2.7 \times 10^3 \text{ kg m}^{-3}$.
- 45 $\frac{F_1}{A_1} = \frac{F_2}{A_2}$ and so $\frac{800}{\pi d^2 / 4} = \frac{1400 \times 9.8}{\pi \times 9.0^2}$ giving $d = 4.3 \text{ m}$.
- 46 Let v_1 be the speed of the water at the faucet of area A_1 . After falling a distance of $h = 5.0 \text{ cm}$ the speed is $v_2 = \sqrt{v_1^2 + 2gh}$ and the cross sectional area is A_2 . The equation of continuity gives $A_1 v_1 = A_2 v_2 = A_2 \sqrt{v_1^2 + 2gh}$. Hence

$$A_1^2 v_1^2 = A_2^2 (v_1^2 + 2gh)$$

$$(A_1^2 - A_2^2) v_1^2 = 2gh A_2^2$$

$$v_1 = \sqrt{\frac{2gh A_2^2}{A_1^2 - A_2^2}}$$

$$v_1 = \sqrt{\frac{2 \times 9.8 \times 0.050 \times 0.60^2}{1.4^2 - 0.60^2}} = 0.470 \text{ m s}^{-1}$$
The flow rate is $Q = A_1 v_1 = 1.4 \times 10^{-4} \times 0.470 = 6.6 \times 10^{-5} \text{ m}^3 \text{ s}^{-1}$.
- 47 The equation of continuity gives $A_1 v_1 = A_2 v_2$ i.e. $\pi \times 0.60^2 \times 1.1 = \pi \times 0.10^2 \times 30 \times v$ and so

$$v = \frac{0.60^2}{30 \times 0.10^2} \times 1.1 = 1.3 \text{ m s}^{-1}.$$
- 48 In one second a mass of water equal to $m = \rho \pi R^2 v = 10^3 \times \pi \times 0.012^2 \times 3.8 = 1.72 \text{ kg}$. The energy of this mass of water as it leaves the pipe is $E = \frac{1}{2} m v^2 + mgh = \frac{1}{2} \times 1.72 \times 3.8^2 + 1.72 \times 9.8 \times 4.0 = 79.8 \approx 80 \text{ J}$. Since this is energy that is provided in 1 s the power is also 80 W.

49 We use Bernoulli's equation to find:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2$$

$$220 \times 10^3 + \frac{1}{2} \times 850 \times 2.0^2 + 0 = p_2 + \frac{1}{2} \times 850 \times 4.0^2 + 850 \times 9.8 \times 8.0$$

$$p_2 = 148.3 \approx 150 \text{ kPa}$$

50 a Imagine a streamline joining the surface to the hole and apply Bernoulli's equation to find:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2$$

$$P_{\text{atm}} + 0 + 1000 \times 9.8 \times 3.0 = P_{\text{atm}} + \frac{1}{2} \times 1000 \times v_2^2 + 0$$

$$v_2 = 7.67 \approx 7.7 \text{ m s}^{-1}$$

b The water from the upper hole has velocity $v = \sqrt{2g \cdot 3.0}$ will hit the ground in a time given by

$$t = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2 \cdot 6.0}{g}}. \text{ It will then cover a horizontal distance of } vt = \sqrt{2g \cdot 3.0} \sqrt{\frac{2 \cdot 6.0}{g}} = 2\sqrt{18} = 8.4853 \text{ m.}$$

The speed from the lower hole will be $v = \sqrt{2g \cdot 6.0}$. The lower hole water will take a time of $t = \sqrt{\frac{2 \cdot 3.0}{g}}$

and cover a distance of $vt = \sqrt{2g \cdot 6.0} \sqrt{\frac{2 \cdot 3.0}{g}} = 2\sqrt{18} = 8.4853 \text{ m}$. The ratio is then 1.

51 The water will exit with speed given by $v = \sqrt{2gd}$. The hole is a distance of $H - d$ from the ground and so

will hit the ground in time $t = \sqrt{\frac{2(H-d)}{g}}$. The range is thus $\sqrt{2gd} \sqrt{\frac{2(H-d)}{g}} = 2\sqrt{d(H-d)}$. This is clearly a

maximum (by graphing or differentiating) when $d = \frac{H}{2}$.

52 a $p_X = P_{\text{atm}} + \rho gh = 1.0 \times 10^5 + 1000 \times 9.8 \times (220 - 95) = 1.325 \times 10^6 \approx 1.3 \times 10^6 \text{ Pa}$;

$$p_Y = P_{\text{atm}} + \rho gh = 1.0 \times 10^5 + 1000 \times 9.8 \times 220 = 2.3 \times 10^6 \text{ Pa.}$$

b i Assuming the surface does not move appreciably, Bernoulli's equation applied to a streamline joining the surface to Y gives

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2$$

$$P_{\text{atm}} + 0 + 1000 \times 9.8 \times 220 = P_{\text{atm}} + \frac{1}{2} \times 1000 \times v_2^2 + 0$$

$$v_2 = 65.66 \approx 66 \text{ m s}^{-1}$$

ii The speed at X can be found by applying the continuity equation from X to Y:

$\pi \times 0.40^2 \times v = \pi \times 0.12^2 \times 65.66$ hence $v = 5.909 \approx 5.9 \text{ m s}^{-1}$. Now applying Bernoulli's equation to a streamline joining the surface to X gives

$$P_{\text{atm}} + 0 + 1000 \times 9.8 \times (220 - 95) = P_X + \frac{1}{2} \times 1000 \times 5.909^2$$

$$P_X = 1.308 \times 10^6 \approx 1.3 \times 10^6 \text{ Pa}$$

The pressure at Y is atmospheric.

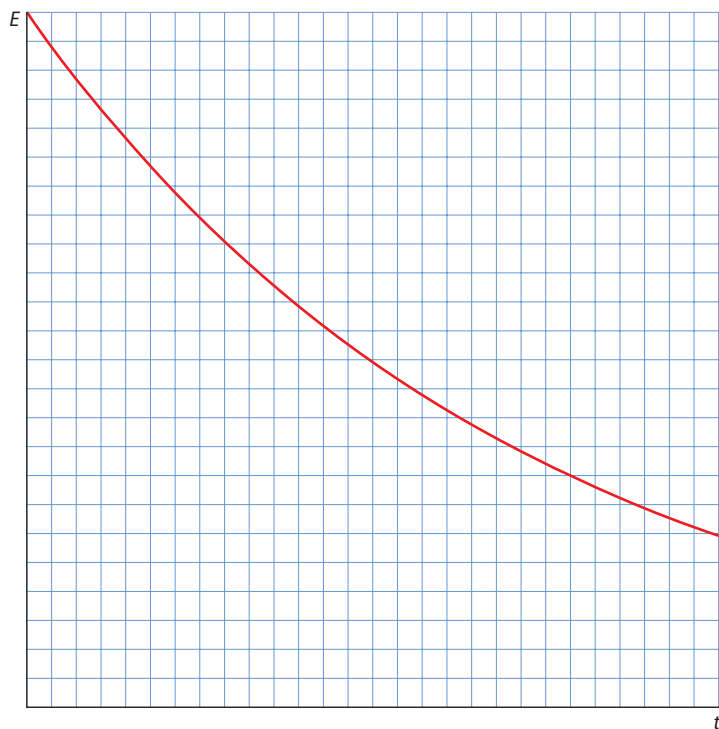
(Comment: the speed at X is not large enough to make much of a difference in pressure between the static case and the case of water flowing.)

- 53 We use Bernoulli's equation $p_1 + \frac{1}{2}\rho v_1^2 + \rho gh_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho gh_2$. Since the heights are approximately the same this becomes $p_1 + \frac{1}{2}\rho v_1^2 = p_2 + \frac{1}{2}\rho v_2^2$ or $p_1 - p_2 = \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2$. From then flow rate $A_1 v_1 = \pi(0.020)^2 v_1 = 1800 \times 10^{-6}$ hence $v_1 = \frac{1800 \times 10^{-6}}{\pi(0.020)^2} = 1.43 \text{ m s}^{-1}$ and $v_2 = 1.43 \times \frac{0.020^2}{0.004^2} = 35.8 \text{ m s}^{-1}$. Hence $p_1 - p_2 = \frac{1}{2} \times 1.20(35.8^2 - 1.43^2) = 398 \text{ Pa}$. Hence $\rho_{\text{Hg}}gh = 398 \text{ Pa} \Rightarrow h = \frac{398}{13600 \times 9.8} = 2.99 \approx 3.0 \text{ mm}$.
- 54 $v = \sqrt{\frac{2\Delta p}{\rho}} = \sqrt{\frac{2 \times 12000}{0.35}} = 261.9 \approx 260 \text{ m s}^{-1}$
- 55 The volume flow rate is $Q = 0.52 = \pi \times 0.40^2 \times v$ and so the speed is $v = \frac{0.52}{\pi \times 0.40^2} = 1.03 \text{ m s}^{-2}$. The Reynolds number is $R = \frac{v\rho r}{\eta} = \frac{1.03 \times 850 \times 0.40}{0.01} = 3.5 \times 10^4$. This is larger than 1000 so the flow is turbulent.
- 56 We assume a wind speed of 10 m s^{-1} a distance between buildings of 10 m an air density of 1 kg m^{-3} and a viscosity of 10^{-5} Pa s . The Reynolds number is $R = \frac{v\rho r}{\eta} = \frac{10 \times 1 \times 5}{10^{-5}} = 5 \times 10^6$, larger than 1000 so turbulent.
- 57 The terminal speed of the droplet is equal to $\frac{2\rho r^2 g}{9\eta} = 4.11 \times 10^{-4}$ and so $r = \sqrt{\frac{9 \times 1.82 \times 10^{-5} \times 4.11 \times 10^{-4}}{2 \times 870 \times 9.8}} = 1.987 \times 10^{-6} \text{ m}$. The mass of the droplet is then $m = \rho V = \rho \frac{4\pi r^3}{3} = 870 \times \frac{4\pi \times (1.987 \times 10^{-6})^3}{3} = 2.8588 \times 10^{-14} \text{ kg}$. When the droplet was balanced in the electric field, $mg = qE$ and so $q = \frac{mg}{E} = \frac{2.8588 \times 10^{-14} \times 9.8}{1.25 \times 10^5} = 2.241 \times 10^{-18} \text{ C}$. Hence $n = \frac{2.241 \times 10^{-18}}{1.6 \times 10^{-19}} = 14.01 \approx 14$.

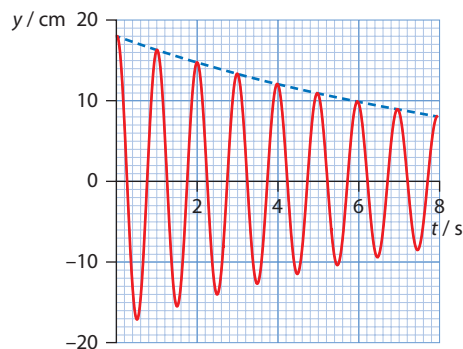
B4 Forced vibrations and resonance (HL)

- 58 Oscillations are called free if no external force acts on the system other than the restoring force that tends to bring the system back to equilibrium. In forced oscillations there is an additional force acting on the system.
- 59 In critically damped oscillations the system, after being displaced, returns to its equilibrium position as fast as possible without performing oscillations. In overdamped oscillations the system will again return to its equilibrium position without oscillations but will do in a long time.
- 60 Damping denotes the loss of energy of an oscillating system which results in a gradual decrease of the amplitude of the oscillations.
- 61 When an oscillating system of natural frequency f_N is exposed to an external periodic force that varies with frequency f_D the system will oscillate with the frequency f_D . The amplitude will be large when $f_D = f_N$ and when this happens we have resonance.

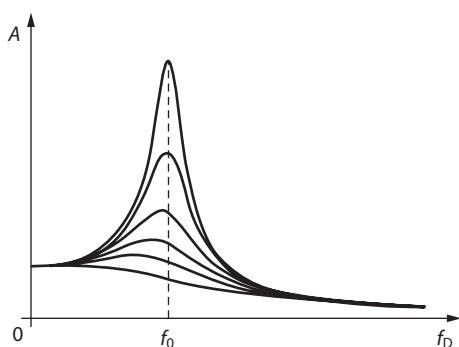
- 62 a The period is 4.0 s and so the frequency is 0.25 Hz.
 b After one oscillation the amplitude drops from 30 cm to 27 cm and so $Q = 2\pi \frac{30^2}{30^2 - 27^2} \approx 33$.
 c The energy decreases exponentially with time:



- d Q will decrease; Q essentially measures how many oscillations the system will perform before coming to rest. With more damping the system will come to rest with fewer oscillations.
 63 a The restoring force must be proportional to and opposite to displacement.
 b See graph shown.

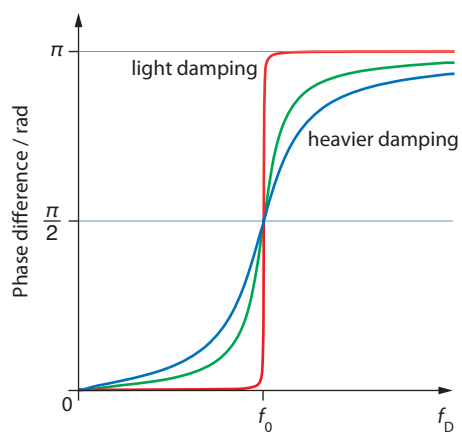


- c i and ii. See graph shown



64 a 10 Hz

b See graph shown.



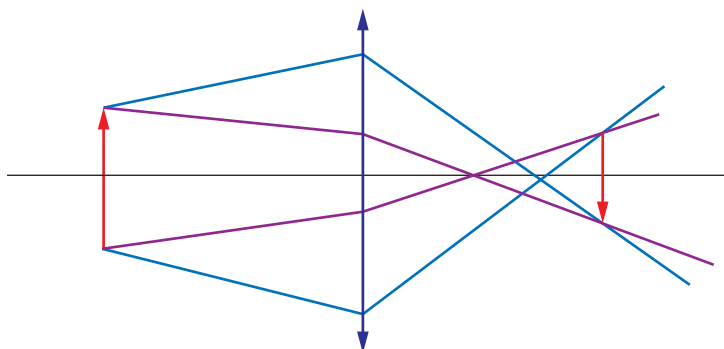
c At 15 Hz the frequency is greater than the natural frequency and so, for light damping, the phase difference between the displacement of the driver and the mass is π . This means that the pendulum mass will move to the left.

Answers to test yourself questions

Option C

C1 Introduction to imaging

- 1 **a** The focal point of a converging lens is that point on the principal axis where a ray parallel to the principal axis refracts through, after passage through the lens.
- b** The focal length is the distance of the focal point from the middle of the lens. In the lens equation this is taken to be a negative number.
- 2 **a** A real image is an image formed by actual rays of light which have refracted through a lens.
- b** A virtual image is formed not by actual rays but by the intersection of their mathematical extensions.
- 3 If a screen is placed at the position of a real image the actual rays of light that go through that image will be reflected off the screen and so the image will be seen on the screen. In the case of a virtual image, placing a screen at the position of the image reveals nothing as there are no rays of light to reflect off the screen.
- 4 No one can explain things better than Feynman and this case is no exception. Search for the YouTube video “Feynman: FUN TO IMAGINE 6: The Mirror” where Feynman explains the apparent left-right case for the mirror. Then try to see what happens with a lens.
- 5 The distance is the focal length so 6.0 cm.
- 6 See graph shown.



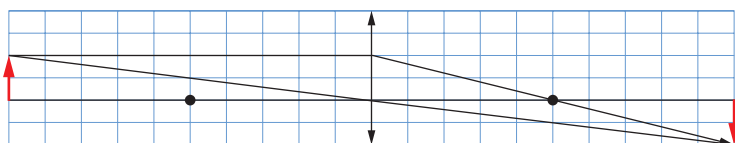
- 7 **a** The diagrams use a vertical scale of 1 cm per line and a horizontal scale of 2 cm per line.

$u = 20$ cm:

The formula gives:

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{10} - \frac{1}{20} = \frac{1}{20} \Rightarrow v = +20 \text{ cm. Further } M = -\frac{v}{u} = -\frac{20}{20} = -1. \text{ So the image is real (positive } v\text{), 20 cm}$$

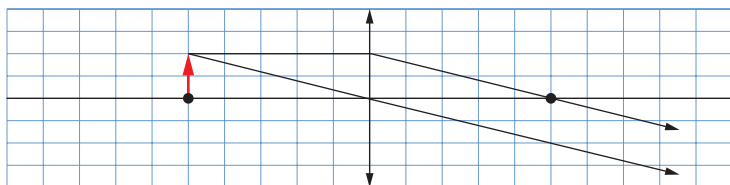
on the other side of the lens, and the image is inverted (negative M) and has height 2 cm ($|M| = 1$). This is what the ray diagram below also gives.



b $u = 10$ cm:

The formula gives:

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{10} - \frac{1}{10} = 0 \Rightarrow v = \infty. \text{ The image is formed at infinity. This is what the ray diagram gives.}$$



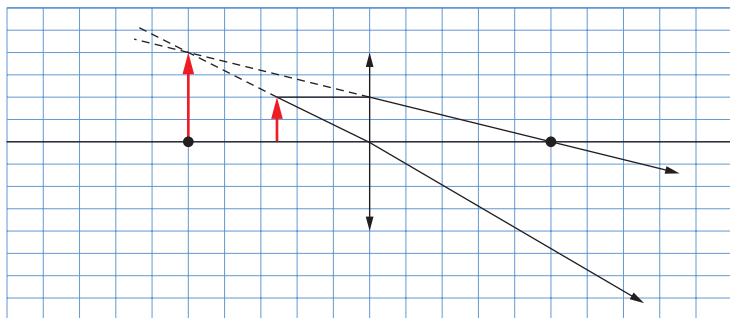
Rays do not meet even when they are extended. Image is said to “form at infinity”.

c $u = 5.0$ cm:

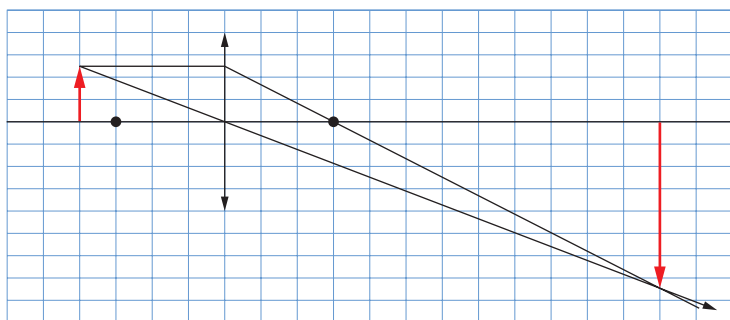
The formula gives:

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{10} - \frac{1}{5.0} = -\frac{1}{10} \Rightarrow v = -10 \text{ cm. Further } M = -\frac{v}{u} = -\frac{-10}{5.0} = +2. \text{ So the image is virtual (negative } v\text{),}$$

10 cm on the same side of the lens, and the image is upright (positive M) and has height twice as large (i.e. 4 cm) (because $|M| = 2$). This is what the ray diagram below also gives.



8 The diagram uses a vertical scale of 1 cm per line and a horizontal scale of 2 cm per line.

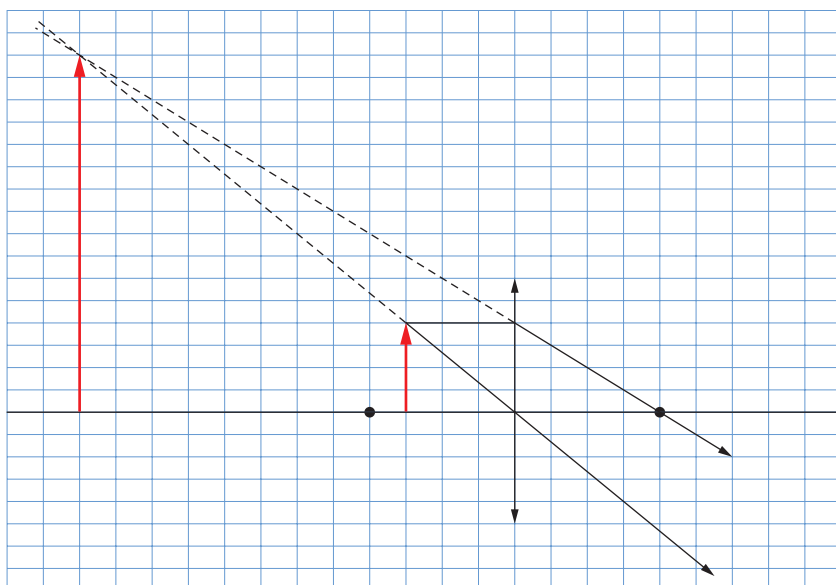


The formula gives:

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{6.0} - \frac{1}{8.0} \Rightarrow v = +24 \text{ cm. Further } M = -\frac{v}{u} = -\frac{24}{8.0} = -3. \text{ So the image is real (negative } v\text{), 24 cm on}$$

the other side of the lens, and the image is inverted (negative M) and has height 3 as large (i.e. 7.5 cm) (because $|M| = 3$). This is what the ray diagram above also gives.

9 See graph shown.



The diagram uses a vertical scale of 1 cm per line and a horizontal scale of 2 cm per line.

The formula gives:

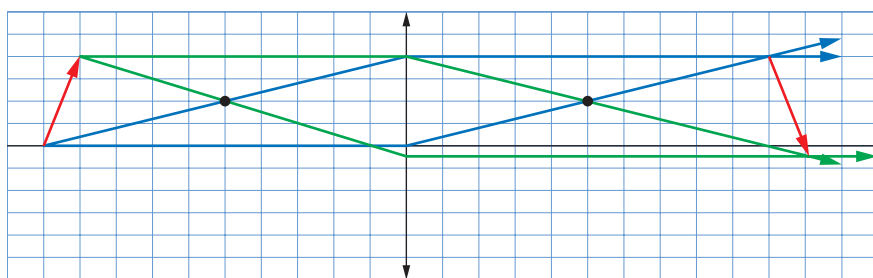
$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{8.0} - \frac{1}{6.0} = -\frac{1}{24} \Rightarrow v = -24 \text{ cm.}$$

Further $M = -\frac{v}{u} = -\frac{-24}{6.0} = +4$. So the image is virtual (negative v), 24 cm on the same side of the lens, and the image is upright (positive M) and has height 4 as large (i.e. 16 cm) (because $|M| = 4$). This is what the ray diagram below also gives.

10 We know that $v > 0$ (real image) and $M = -1 = -\frac{v}{u}$ (same size). Hence $v = u$ and so

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Rightarrow \frac{2}{u} = \frac{1}{f} \Rightarrow u = 2f = 9.0 \text{ cm.}$$

11 See the diagram that follows. The rays from the top of the object have been drawn green and those from the bottom blue for the sake of clarity.



The top of the image will be formed at a distance from the lens given by:

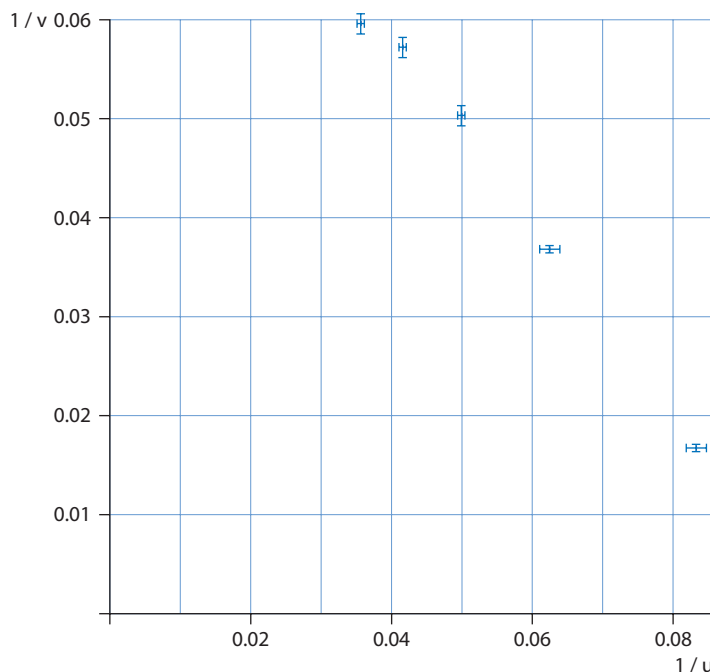
$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{5.0} - \frac{1}{9.0} \Rightarrow v = 11.25 \text{ cm}$$

$$\text{and the bottom at a distance of } \frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{5.0} - \frac{1}{10.0} \Rightarrow v = 10.0 \text{ cm.}$$

The angle the object makes with the horizontal is $\tan^{-1} \frac{4}{1} \approx 76^\circ$. The image makes an angle given by $\tan^{-1} \frac{4.5}{1.25} \approx 75^\circ$ so it is slightly smaller.

- 12 a Since we know that $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$ we should plot $\frac{1}{u}$ versus $\frac{1}{v}$. We expect a straight line with gradient equal to -1 and equal vertical and horizontal intercepts equal to $\frac{1}{f}$.

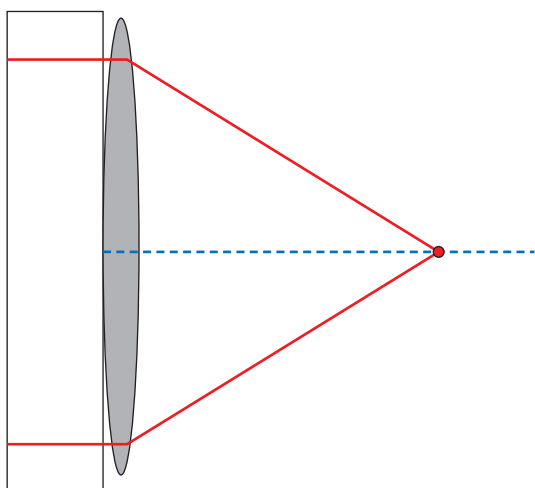
b The graph shows the data points and the (small) error bars.



The line of best fit is $\frac{1}{v} = 0.096 - \frac{0.95}{u}$. The intercepts are 0.096 and $\frac{0.096}{0.95}$ giving focal lengths

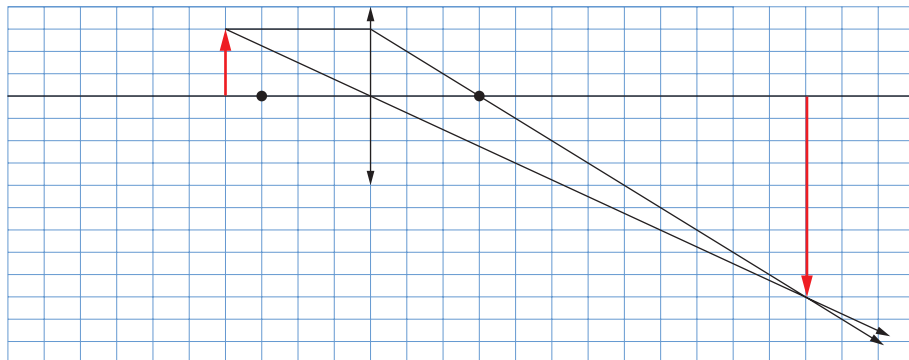
$\frac{1}{f} = 0.096 \Rightarrow f = 10.42 \text{ cm}$ and $\frac{1}{f} = \frac{0.096}{0.95} \Rightarrow f = 9.896 \text{ cm}$. So approximately, $f = 10.2 \pm 0.3 \text{ cm}$.

- 13 From the diagram it should be clear that rays must hit the mirror at right angles if they are to return to the position of the object. This means that the distance of the object from the lens when the object and image coincide is the focal length.



- 14 a $\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{15} - \frac{1}{20} \Rightarrow v = +60$ cm. Further $M = -\frac{v}{u} = -\frac{60}{20} = -3$. So the image is real (positive v), 60 cm on the other side of the lens, and the image is inverted (negative M) and has height 3 times as large (because $|M| = 3$). This is what the ray diagram below also gives.

b See graph shown.

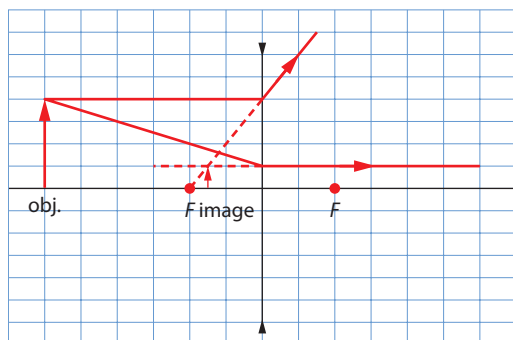


- 15 a We must have that $u + v = 5$ and $\frac{1}{u} + \frac{1}{v} = \frac{1}{0.6}$ where distances are in meters. Then $v = 50 - u$ and so $\frac{1}{5-v} + \frac{1}{v} = \frac{1}{0.6}$. This gives $v^2 - 5v + 3 = 0$ with solutions $v = 4.30$ m or $v = 0.70$ m.

b In the first case the magnification is $M = -\frac{4.30}{0.70} = -6.1$ and in the second $M = -\frac{0.70}{4.30} = -0.16$ so the magnification is larger (in magnitude) in the first case.

- 16 We use $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ with $f = -4.0$ cm. Hence $\frac{1}{v} = -\frac{1}{4.0} - \frac{1}{12} = -\frac{1}{3.0}$. Hence $v = -3.0$ cm and $M = -\frac{-3.0}{12} = +\frac{1}{4.0}$.

The image is virtual ($v < 0$), upright and smaller by a factor of 4 ($M = +\frac{1}{4.0}$).



- 17 Let u be the distance of an object from the first lens. The lens creates an image a distance v from the lens which

is given by $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ hence $\frac{1}{v} = \frac{1}{f_1} - \frac{1}{u}$. This image serves as the virtual object in the second lens. Because

the lenses are thin the distance of this virtual object is also v . Hence the final image is formed at distance

of $-\left(\frac{1}{f_1} - \frac{1}{u}\right) + \frac{1}{v_2} = \frac{1}{f_2}$ (the minus sign in front of the first term is because the object is virtual) and so

$\frac{1}{v_2} = \left(\frac{1}{f_1} - \frac{1}{u}\right) + \frac{1}{f_2}$. This is what we would have obtained if the inverse focal length of the combination were

$$\frac{1}{f} = \frac{1}{f_2} + \frac{1}{f_1}. \text{ Hence } f = \frac{f_1 f_2}{f_1 + f_2}.$$

18 Using $f = \frac{f_1 f_2}{f_1 + f_2}$ we find $f = \frac{10.0 \times 4.0}{14.0} = 2.86 \text{ cm}$.

19 a The image in the first lens is found from: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Rightarrow \frac{1}{40.0} + \frac{1}{v} = \frac{1}{15.0} \Rightarrow v = 24.0 \text{ cm}$. This means that the distance of the image from the second lens is 1.0 cm . This image now serves as the object for the second lens. So $\frac{1}{1.0} + \frac{1}{v} = \frac{1}{2.0} \Rightarrow v = -2.0 \text{ cm}$. The final image is 2.0 cm to the left of the second lens.

b The overall magnification is the product of the individual lens magnifications i.e.

$$M = M_1 M_2 = \left(-\frac{24}{40}\right) \left(-\frac{2.0}{1.0}\right) = -1.2.$$

c Since $M < 0$, the final image is inverted relative to the original object and is 1.2 times larger.

20 a The image in the first lens is found from: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Rightarrow \frac{1}{30.0} + \frac{1}{v} = \frac{1}{35.0} \Rightarrow v = -210 \text{ cm}$. This means that the distance of the image from the second lens is 235 cm . This image now serves as the object for the second lens. So $\frac{1}{235} + \frac{1}{v} = -\frac{1}{20.0} \Rightarrow v = -18.4 \text{ cm}$. The final image is 18.4 cm to the left of the second lens.

b The overall magnification is the product of the individual lens magnifications i.e.

$$M = M_1 M_2 = \left(-\frac{-210}{30}\right) \left(-\frac{-18.4}{235}\right) = +0.548 \approx 0.55.$$

c Since $M > 0$, the final image is upright relative to the original object and is 0.55 times smaller.

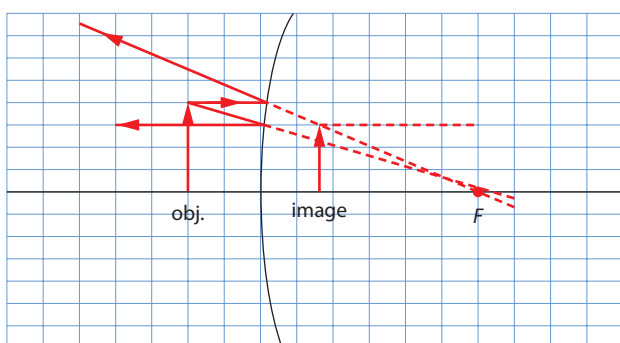
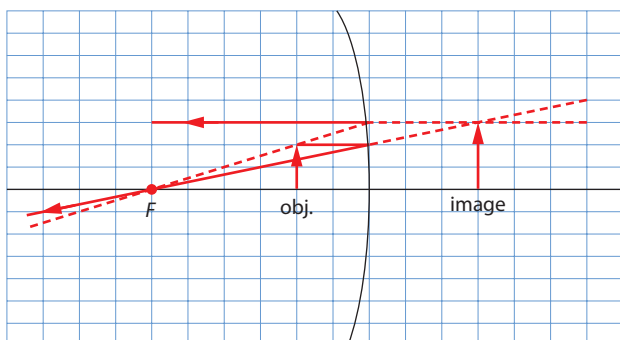
21 a The mirror formula gives $\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Rightarrow \frac{1}{v} = \frac{1}{12} - \frac{1}{4.0} \Rightarrow v = -6.0 \text{ cm}$. The magnification is $M = -\frac{-6.0}{4.0} = +1.5$.

Therefore the image is virtual, formed on the same side of the mirror as the object, upright and 1.5 times taller.

b Now the mirror formula gives $\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Rightarrow \frac{1}{v} = -\frac{1}{12} - \frac{1}{4.0} \Rightarrow v = -3.0 \text{ cm}$. The magnification is

$M = -\frac{-3.0}{4.0} = +0.75$. Therefore the image is virtual, formed on the same side of the mirror as the object, upright and 0.75 times the object's height.

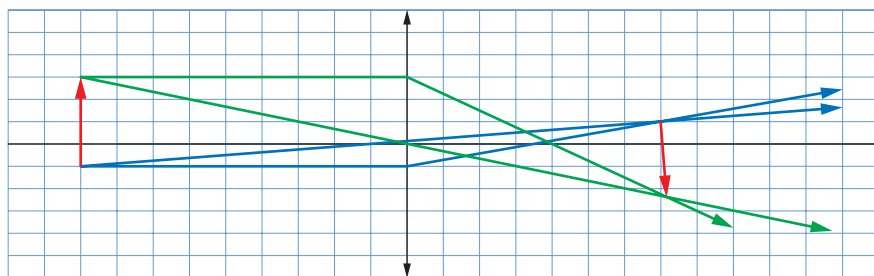
c See graphs shown.



22 We are told that $M = +2.0$ (image is upright so $M > 0$) hence $2.0 = -\frac{v}{u} \Rightarrow v = -2.0u = -24$ cm. Hence from the mirror formula gives $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ we get $\frac{1}{f} = \frac{1}{12} - \frac{1}{24} \Rightarrow f = 24$ cm. Since the focal length is positive the mirror is concave.

23 a There are two main lens aberrations. In spherical aberration rays that are far from the principal axis have a different focal length than rays close to the principal axis. This results in images that are blurred and curved at the edges. In chromatic aberration, rays of different wavelength have slightly different focal lengths resulting in images that are blurred and coloured. Spherical aberration is reduced by only allowing rays close to the principal axis to enter the lens and chromatic aberration is reduced by combining the lens with a second diverging lens.

b i The diagram shows (under the simple conditions of this problem) that a different focal length (depending on the distance of the paraxial rays from the principal axis) creates an image that is curved at the edges.



ii The image drawn with one focal length would be straight.

24 We are told that $u = f + x$ and $v = f + y$. Then $\frac{1}{f+x} + \frac{1}{f+y} = \frac{1}{f} \Rightarrow \frac{(f+x)(f+y)}{f+x+f+y} = f$. Simplifying,

$$f^2 + fx + fy + xy = f(2f + x + y) = 2f^2 + fx + fy$$

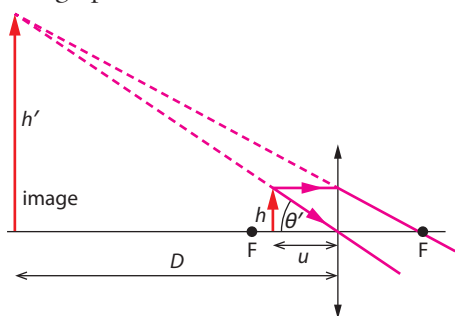
$$xy = f^2$$

25 a The image is virtual so $v = -25$ cm. $\frac{1}{u} + \frac{1}{v} = \frac{1}{10.0} \Rightarrow \frac{1}{u} - \frac{1}{25} = \frac{1}{10.0} \Rightarrow u = 7.143 \approx 7.1$ cm.

b At the focal point of the lens, 10 cm away.

c The angular magnification in this case is $M = \frac{25}{f} = \frac{25}{10} = 2.5$ and $M = \frac{\theta'}{\theta}$. Now $\theta \approx \frac{1.6 \times 10^{-3}}{0.25} = 0.0064$ rad and so $\theta' \approx 2.5 \times 0.0064 = 0.016$ rad.

26 a See graph shown.



b The nearest point to the eye where the eye can focus without straining.

c When the image is formed at the near point (25 cm away) we have that $v = -25$ cm.

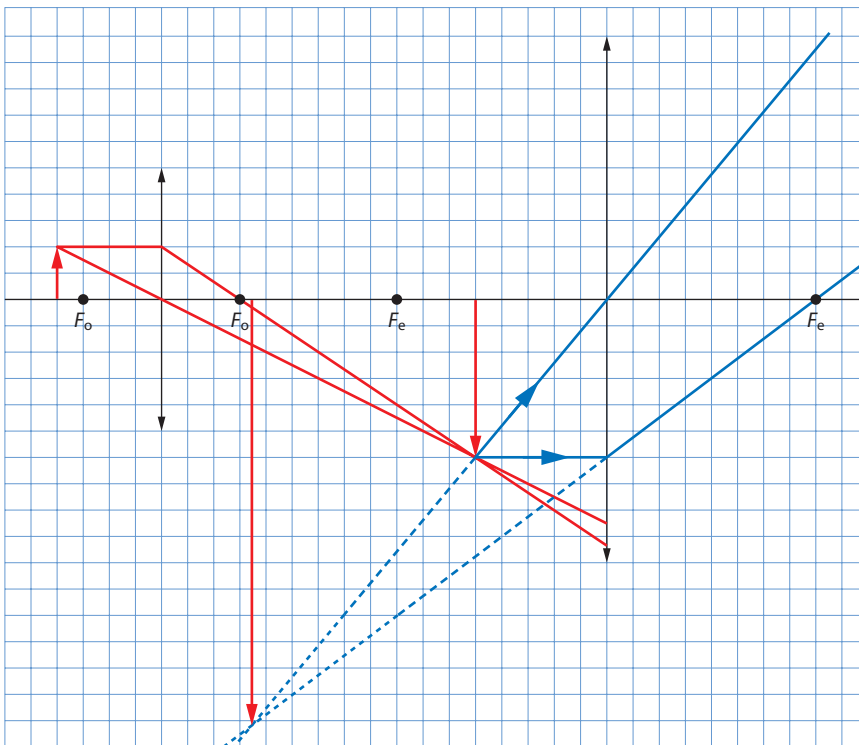
Hence $\frac{1}{u} + \frac{1}{-25} = \frac{1}{f} \Rightarrow \frac{1}{u} = \frac{1}{f} + \frac{1}{25} = \frac{25+f}{25f} \Rightarrow u = \frac{25f}{25+f}$. The magnification is then

$$M = -\frac{v}{u} = -\frac{-25}{\frac{25f}{25+f}} = +\frac{25+f}{f} = 1 + \frac{25}{f}$$

- 27 The magnification is $M = 1 + \frac{25}{f} = 1 + \frac{25}{5.0} = 6.0$. Hence if d is the distance of the two points we must have that $0.12 \times 10^{-3} = 6.0d$ and so $d = 2.0 \times 10^{-5}$ m.

C2 Imaging instrumentation

- 28 a The image in the objective is formed at a distance given by: $\frac{1}{1.50} + \frac{1}{v} = \frac{1}{0.80} \Rightarrow v = 1.71 \approx 1.7$ cm. The magnification of the objective is then $m_0 = -\frac{1.71}{1.50} = -1.14$.
- b $\frac{1}{u} + \frac{1}{-25} = \frac{1}{4.0} \Rightarrow u = 3.45 \approx 3.4$ cm
- c The magnification of the eyepiece is $1 + \frac{25}{4.0} = 7.25$. The overall magnification is then $-1.14 \times 7.25 = -8.3$.
- 29 a From $\frac{1}{25} + \frac{1}{v_1} = \frac{1}{20}$ we get $v_1 = 100$ mm.
- b From $\frac{1}{u_2} - \frac{1}{350} = \frac{1}{80}$ we get $u_2 = 65.12 \approx 65$ mm.
- c The overall magnification is $M = m_o \times m_e \times \frac{D}{v_2} = \left(-\frac{100}{25}\right) \times \left(-\frac{350}{65.12}\right) \times \frac{250}{350} = -15.4 \approx -15$.
- 30 See diagram below.



- 31 The objective forms the first image at a distance v_1 from the objective where $\frac{1}{30} + \frac{1}{v_1} = \frac{1}{24}$ and so $v_1 = 120$ mm. The magnification of the objective is $m_o = -\frac{120}{30} = -4.0$. The overall magnification is $M = m_o \times m_e = m_o \times \left(1 + \frac{D}{f_e}\right)$ i.e. $-30 = -4.0 \times \left(1 + \frac{250}{f_e}\right)$ and so $1 + \frac{250}{f_e} = 7.5$ giving $f_e = 38.5 \approx 38$ mm.
- 32 The blue line through the middle of the eyepiece lens is a construction ray.

- 33 a The final image is formed at infinity.
 b $M = \frac{f_o}{f_e} \Rightarrow 14 = \frac{2.0}{f_e}$. Hence $f_e = 0.14$ m.
- 34 a The angular width of the moon is $\theta = \frac{3.5 \times 10^6}{3.8 \times 10^8} = 0.00921 \approx 0.0092$ rad. (This is
 $\theta = 0.00921 \times \frac{180^\circ}{\pi} = 0.527^\circ \approx 0.53^\circ$.)
 b The angular magnification is $M = \frac{f_o}{f_e} = \frac{3.6}{0.12} = 30$. The diameter of the image of the moon is then
 $30 \times 0.00921 = 0.276 \approx 0.28$ rad.
- 35 a The angular magnification is $M = \frac{f_o}{f_e} = \frac{80}{20} = 4.0$.
 b The angle subtended by the building without a telescope is $\theta = \frac{65}{2500} = 0.0260$ rad and so the angle subtended
 by the image is $\theta' = M\theta = 4 \times 0.0260 = 0.104$ rad.
- 36 a $M = \frac{f_o}{f_e} = \frac{67}{3.0} = 22.3 \approx 22$.
 b $f_o + f_e = 70$ cm
- 37 The objective focal length must be 57 cm. If the final image is formed at infinity, it means that the image in the
 objective is formed at a distance of 3.0 cm from the eyepiece i.e. $61.5 - 3.0 = 58.5$ cm from the objective lens.
 Hence $\frac{1}{u} + \frac{1}{58.5} = \frac{1}{57.0} \Rightarrow u = 2223$ cm ≈ 22 m.
- 38 a A technique in radio astronomy in which radio waves emitted by distant sources are observed by an array of
 radio telescopes which combine the individual signals into one.
 b $\theta \approx 1.22 \times \frac{\lambda}{b} = 1.22 \times \frac{0.21}{25 \times 10^3} = 1.0 \times 10^{-5}$ rad.
 c The smallest angular separation that can be resolved is 1.0×10^{-5} rad. The smallest distance is therefore
 $2 \times 10^{22} \times 1.0 \times 10^{-5} = 2 \times 10^{17}$ m.
- 39 The universe is full of sources that emit at all parts of the electromagnetic spectrum not just optical light.
- 40 They do not suffer from spherical aberrations.
- 41 Advantages: free of atmospheric turbulence and light pollution; no atmosphere to absorb specific wavelengths
 Disadvantages: expensive to put in orbit; expensive to service.

C3 Fibre optics

- 42 $n = \frac{c}{c_m} \Rightarrow c_m = \frac{c}{n} = \frac{3 \times 10^8}{1.45} = 2.07 \times 10^8$ m s⁻¹.
- 43 a The phenomenon in which a ray approaching the boundary of two media reflects without any refraction
 taking place.
 b Critical angle is that angle of incidence for which the angle of refraction is 90°.
 c The critical angle is found from $n_1 \sin \theta_c = n_2 \sin 90^\circ \Rightarrow \sin \theta_c = \frac{n_2}{n_1}$. Since the sine of an angle cannot exceed 1
 we must have $n_2 < n_1$ for the critical angle to exist. So total internal reflection is a one way phenomenon.
- 44 $n_1 \sin \theta_c = n_2 \sin 90^\circ \Rightarrow \sin \theta_c = \frac{n_2}{n_1} = \frac{1.46}{1.50} \Rightarrow \theta_c = 76.7^\circ$.

45 a We know that: $n_1 \sin \theta_C = n_2 \sin 90^\circ$ hence $\sin \theta_C = \frac{n_2}{n_1}$.

$$\text{Then } \cos \theta_C = \sqrt{1 - \sin^2 \theta_C} = \sqrt{1 - \frac{n_2^2}{n_1^2}} = \frac{\sqrt{n_1^2 - n_2^2}}{n_1}.$$

b From the diagram, $1 \sin A = n_1 \sin a$. But $a = 90^\circ - \theta_C$ and so $\sin a = \sin(90^\circ - \theta_C) = \cos \theta_C$. Hence

$$\sin A = n_1 \cos \theta_C = n_1 \times \frac{\sqrt{n_1^2 - n_2^2}}{n_1} = \sqrt{n_1^2 - n_2^2}.$$

c $A = \arcsin \sqrt{n_1^2 - n_2^2} = \arcsin \sqrt{1.50^2 - 1.40^2} = 32.6^\circ$.

46 $A = \arcsin \sqrt{n_1^2 - n_2^2} = \arcsin \sqrt{1.52^2 - 1.44^2} = 29.1^\circ$.

47 We must have $\sqrt{n_1^2 - 1.42^2} = 1 \Rightarrow n_1 = 1.7367 \approx 1.74$.

48 It has to be exceptionally pure.

49 a Dispersion is the phenomenon in which the speed of a wave depends on wavelength. This means that the different wavelength components of a beam of light will take different times to travel the same distance.

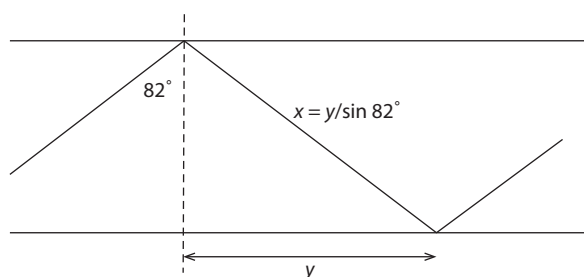
b Material dispersion is the dispersion discussed in (a). Waveguide dispersion has to do with rays of light following different paths in an optic fibre and hence taking different times to arrive at their destination.

50 a $n = \frac{c}{c_m} \Rightarrow c_m = \frac{c}{n} = \frac{3 \times 10^8}{1.52} = 1.9737 \times 10^8 \approx 1.97 \times 10^8 \text{ m s}^{-1}$.

b The shortest time will be for a ray that travels down the length of the fibre on a straight line of length 8.0 km, i.e. the time of travel will be $\frac{8.0 \times 10^3}{1.9737 \times 10^8} = 4.05 \times 10^{-5} \text{ s}$. The longest time of travel will be for

that ray that undergoes as many internal reflections as possible. The length of the path travelled is then

$$\frac{8.0}{\sin 82^\circ} = 8.0786 \approx 8.08 \text{ km (see diagram) and so the time is } \frac{8.0786 \times 10^3}{1.9737 \times 10^8} = 4.09 \times 10^{-5} \text{ s}.$$



51 The height of the pulses will be less and the width of the pulses greater.

52 a A monomode optic fibre is a fibre with a very thin core so that effectively all rays entering the fibre follow the same path. In a multimode fibre (which is thicker than a monomode fibre) rays follow very many paths of different length in getting to their destination.

b The transition from multimode to monomode fibres offers a very large increase in bandwidth. As discussed also in question 13, dispersion limits the maximum frequency that can be transmitted and hence the bandwidth. A very small diameter monomode fibre will suffer the least from modal dispersion (and hence the distortion and widening of the pulse) and material dispersion is also minimised by using lasers rather than LED's. Hence the bandwidth is increased as the monomode fibre diameter is decreased and laser light is used.

53 Advantages include:

- (i) the low attenuation per unit length which means that a signal can travel large distances before amplification
- (ii) increased security because the signal can be encrypted and the transmission line itself cannot easily be tampered with
- (iii) large bandwidth and so a large information carrying capacity
- (iv) not susceptible to noise
- (v) they are thin and light and
- (vi) do not radiate so there is no crosstalk between lines that are close to each other.

54 The main cause of attenuation in an optic fibre is scattering of light off impurities in the glass making up the core of the fibre.

55 Let P_{in} be the power in to the first amplifier. Then the power out of the first amplifier is $P' = P_{\text{in}} \times 10^{\frac{G_1}{10}}$. This is

$$\text{input to the second amplifier so its output is } P_{\text{out}} = \left(P_{\text{in}} \times 10^{\frac{G_1}{10}} \right) \times 10^{\frac{G_2}{10}} = P_{\text{in}} \times 10^{\frac{G_1 + G_2}{10}} = P_{\text{in}} \times 10^{\frac{G_1 + G_2}{10}} \text{ showing}$$

that the gain overall is $G_1 + G_2$.

56 The power loss is $10 \log \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \log \frac{3.20}{4.60} = -1.58 \text{ dB}$.

57 The power loss is $10 \log \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \log \frac{5.10}{8.40} = -2.167 \text{ dB}$. So the loss per km is $\frac{2.167}{25} = 0.087 \text{ dB km}^{-1}$.

58 The power loss when the power falls to 70% of the original input power is

$$10 \log \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \log \frac{0.70P}{P} = -1.55 \text{ dB. So, } 12 \times L = 1.55 \Rightarrow L = 0.13 \text{ km.}$$

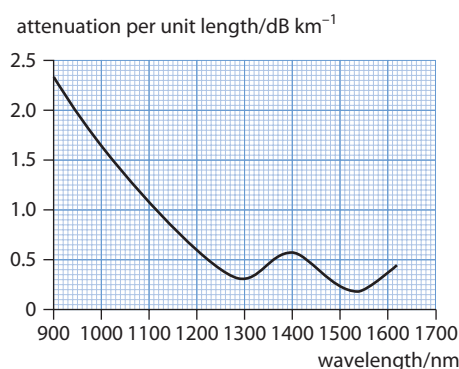
59 There is no overall gain in power since $+15 - 12 = 3.0 \text{ dB}$. Let the input power be P . Then the output power is

$$P' = P \times 10^{\frac{3}{10}} = P \times 10^{0.3} \approx 2.0P.$$

60 There is no overall gain or loss in power since $+7 - 10 + 3 = 0 \text{ dB}$. So the output power is the same as the input power, the ratio is 1.

61 The overall gain is $10 \log \frac{P_{\text{out}}}{P_{\text{in}}} = 10 \log \frac{2P}{P} = 10 \log 2 \approx 3.0 \text{ dB}$. Hence $-12 + G - 6.0 = 3.0 \text{ dB}$ giving $G = 21 \text{ dB}$.

62 a See graph.



b The attenuation per unit length is least for long wavelengths, in particular 1310 nm and 1550 nm, and these are infrared wavelengths.

C4 Medical Imaging

- 63 **a** Attenuation is the loss of energy in a beam as it travels through a medium.
b For X-rays the main mechanism for attenuation is the photoelectric effect in which X-ray photons knock electrons off the medium's atoms, losing energy in the process. This effect is dependent on the medium atoms' atomic number Z . This means that media with different Z have different attenuation which allows an image of the boundary of the two media to be made.
- 64 The image can be formed faster by using intensifying screen. This is done by having X-rays that have gone through the patient strike a screen containing fluorescent crystals which then emit visible light. The visible light helps form the image on photographic film faster.
- 65 If neighbouring organs have the same atomic number the boundary of the organs will not appear clearly. By having the patient swallow a barium meal, the atomic number of organs such as the stomach or the intestine is now greater and can be distinguished from its surrounding tissue.
- 66 The blurry images are caused by X-rays that have scattered in the patient's body and thus now deviate from their original directions. This may be minimised by placing lead strips between the patient and the film, along the direction of the incident X-ray beam. In this way scattered X-rays will be blocked by the strip and not fall on the film.
- 67 **a** For the top curve the HVT is 6.0 mm and for the other it is about 4.0 mm.
b The larger energy corresponds to the curve with the longer HVT.
- 68 **a** The HVT is about 5.0 mm and so the linear attenuation coefficient is about $\frac{\ln 2}{5.0} = 0.139 \approx 0.14 \text{ mm}^{-1}$.
b The transmitted intensity must be 20% of the incident and from the graph this corresponds to a length of about 11.5 mm.
- 69 $0.60 = e^{-\mu \times 4}$ and so $\mu = -\frac{1}{4} \ln 0.6 = 0.128 \text{ mm}^{-1}$. Then, $0.20 = e^{-\mu x}$ and so

$$x = -\frac{1}{\mu} \ln 0.2 = -\frac{1}{0.128} \ln 0.2 = 12.6 \approx 13 \text{ mm}.$$
- 70 $\mu = \frac{1}{3} \ln 2 = 0.231 \text{ mm}^{-1}$ and so $I = I_0 e^{-0.231 \times 1} = 0.794 I_0 \approx 0.8 I_0$.
- 71 It means that as the beam moves through the metal the proportion of the total energy of the X-rays carried by high energy photons increases. This is because the low energy photons get absorbed leaving only the high energy photons move through. For the 20 keV photons the transmitted intensity is $I_{20} = I_0 e^{-\frac{\ln 2}{2.2} \times 5} = 0.207 I_0$. For the 25 keV photons it is $I_{25} = I_0 e^{-\frac{\ln 2}{2.8} \times 5} = 0.290 I_0$. Hence $\frac{I_{25}}{I_{20}} = \frac{0.290 I_0}{0.207 I_0} = 1.4$.
- 72 Ultrasound is sound of frequency higher than about 20 kHz. It is produced by applying an alternating voltage to certain crystals which vibrate as a result emitting ultrasound.
- 73 The wavelength of this ultrasound is $\lambda = \frac{v}{f} = \frac{1540}{5 \times 10^6} = 3.1 \times 10^{-4} \text{ m} = 0.3 \text{ mm}$. The order of magnitude of the size that can be resolved is of the order of the wavelength and so about 0.3 mm.
- 74 **a** Impedance is the product of the density of a medium and the speed of sound in that medium.
b $Z = \rho c \Rightarrow c = \frac{Z}{\rho} = \frac{1.4 \times 10^6}{940} = 1.5 \times 10^3 \text{ m s}^{-1}$.
- 75 **a** The fraction of the transmitted intensity is given by $\frac{4Z_1 Z_2}{(Z_1 + Z_2)^2}$ and in this case equals

$$\frac{4 \times 420 \times 1.6 \times 10^6}{(420 + 1.6 \times 10^6)^2} \approx 1.0 \times 10^{-3}.$$

b This is a very small fraction of the incident intensity and not enough to be useful for diagnostic purposes. More intensity has to be transmitted which is why the gel like substance is put in between the skin and the transducer; the gel has an intensity closer to the tissue's so more intensity gets transmitted.

- 76 In the A the signal strength may be converted to a dot whose brightness is proportional to the signal strength. We now imagine a series of transducers along an area of the body. Putting together the images (as dots) of each transducer forms a **two-dimensional image** of the surfaces that cause reflection of the ultrasound pulses. This creates a B scan.
- 77 The difference in energy for spin up and down states depends on the magnetic field strength; only those protons in regions where the magnetic field has the “right” value will be absorbed and so it is possible to determine where these photons have been emitted from. This is achieved by exposing the patient to an additional non-uniform field, the gradient field.
- 78 In this imaging technique, the patient is not exposed to any harmful radiation.

The method is based on the fact that protons have a property called spin. The proton’s spin will align parallel or anti-parallel to a magnetic field and the energy of the proton will depend on whether its spin is **up** (i.e. parallel to the magnetic field) or **down** (opposite to the field). The state with spin up has a lower energy than that of spin down. The **difference** in energy depends on the magnetic field strength.

The patient is placed in an enclosure that creates a very **uniform** magnetic field throughout the body. A source of radio frequency forces protons with spin up to make a transition to a state with spin down. As soon as this happens the protons will make a transition back down to the spin up state emitting a photon in the process.

The idea is to detect these photons and correlate them with the point from which they were emitted. This is done with the help of a gradient field as discussed in the previous question.

Answers to test yourself questions

Option D

D1 Stellar quantities

1 The distance is $\frac{1}{0.285} = 3.51$ pc.

2 The distance is $\frac{10.8}{3.26} = 3.31$ pc and so the parallax angle is $\frac{1}{3.31} = 0.302''$.

3 a The distance is $\frac{1}{0.0067} = 149$ pc.

b The diameter is

$$D = d\theta = 149 \times \frac{0.016}{3600} \times \frac{\pi}{180} \text{ pc} = 149 \times \frac{0.016}{3600} \times \frac{\pi}{180} \times 3.26 \times 9.46 \times 10^{15} = 3.56 \times 10^{11} \text{ m. The radius is}$$

$$\frac{3.56 \times 10^{11}}{2} = 1.78 \times 10^{11} \text{ m. This is about 256 times larger than the radius of the sun.}$$

4 From Topic 12 we know that the nuclear radius for a nucleus of mass number A is given by $1.2 \times A^{1/3} \times 10^{-15}$ m. The nuclear volume is then

$$\begin{aligned} V &= \frac{4\pi}{3} R^3 \\ &= \frac{4\pi}{3} (1.2 \times A^{1/3} \times 10^{-15})^3 \\ &= 7.24 \times 10^{-45} \times A \text{ m}^3 \end{aligned}$$

The mass of the nucleus is A u, i.e. $A \times 1.66 \times 10^{-27}$ kg. The density is therefore (note how A cancels out)

$$\begin{aligned} \rho &= \frac{A \times 1.66 \times 10^{-27}}{7.24 \times 10^{-45} \times A} \\ &\approx 2.3 \times 10^{17} \text{ kg m}^{-3} \end{aligned}$$

So all nuclei have the same density and that density is comparable to a neutron star density.

5 The diameter d will be approximately equal to $d = D\theta$ where $D = 1.5 \times 10^{11}$ m is the earth – sun distance and θ is the angle subtended by the sunspot in radians. Hence, $d = 1.5 \times 10^{11} \times \frac{4}{3600} \times \frac{\pi}{180} = 3 \times 10^6$ m.

6 The diameter d will be approximately equal to $d = D\theta$ where $D = 3.8 \times 10^8$ m is the earth – moon distance and θ is the angle subtended in radians. Hence, $d = 3.8 \times 10^8 \times \frac{0.05}{3600} \times \frac{\pi}{180} = 92 \approx 10^2$ m.

7 The speed of the sun is $v = \frac{2\pi r}{T} = \frac{2\pi \times 2.8 \times 10^4 \times 9.46 \times 10^{15}}{211 \times 10^6 \times 365 \times 24 \times 3600} \approx 2.5 \times 10^5 \text{ m s}^{-1}$. Now using

$$v^2 = \frac{GM}{r} \Rightarrow M = \frac{v^2 r}{G}, \text{ we find } M = \frac{(2.5 \times 10^5)^2 \times 2.8 \times 10^4 \times 9.46 \times 10^{15}}{6.67 \times 10^{-11}} = 2.5 \times 10^{41} \text{ kg. This is the mass that is enclosed within a radius of 28000 ly. The mass in the galaxy outside this radius does not influence the motion of the sun.}$$

8 a See discussion in textbook.

b The method fails for stars far away (more than about 300 pc or 1000 ly) because then the parallax angle is too small to be measured accurately.

9 We use $b = \frac{L}{4\pi d^2} \Rightarrow L = b \times 4\pi d^2$ so that $L = 3.0 \times 10^{-8} \times 4\pi \times (70 \times 9.46 \times 10^{15})^2$ i.e. $L = 1.7 \times 10^{29}$ W.

10 We use $b = \frac{L}{4\pi d^2}$ so that $b = \frac{4.5 \times 10^{28}}{4\pi \times (88 \times 9.46 \times 10^{15})^2} = 5.2 \times 10^{-9}$ W m⁻².

11 From $b = \frac{L}{4\pi d^2}$ we get $d = \sqrt{\frac{L}{4\pi b}}$, i.e. $d = \sqrt{\frac{6.2 \times 10^{32}}{4\pi \times 8.4 \times 10^{-10}}} = 2.4 \times 10^{20}$ m. This corresponds to $\frac{2.4 \times 10^{20}}{9.46 \times 10^{15}} \approx 26$ kly.

12 a From $L = \sigma AT^4$, $\frac{L_H}{L_C} = \frac{\sigma A(4T)^4}{\sigma AT^4} = 4^4 = 256$.

$$\text{b } \frac{b_H}{b_C} = \frac{\left(\frac{L_H}{4\pi d_H^2}\right)}{\left(\frac{L_C}{4\pi d_C^2}\right)} \Rightarrow 1 = \frac{L_H}{L_C} \times \frac{d_C^2}{d_H^2} = 256 \frac{d_C^2}{d_H^2} \text{ and so } \frac{d_C^2}{d_H^2} = \frac{1}{256} \Rightarrow \frac{d_C}{d_H} = \frac{1}{16}$$

13 $\frac{b_A}{b_B} = \frac{\left(\frac{L_A}{4\pi d^2}\right)}{\left(\frac{L_B}{4\pi d^2}\right)} \Rightarrow \frac{9.0 \times 10^{-12}}{3.0 \times 10^{-13}} = \frac{L_A}{L_B}$, i.e. $\frac{L_A}{L_B} = 30$.

14 a Since $L = \sigma AT^4 = \sigma 4\pi R^2 T^4$: (a) $1 = \frac{R_A^2 (5000)^4}{R_B^2 (10000)^4} \Rightarrow \frac{R_A}{R_B} = \sqrt{\frac{(10000)^4}{(5000)^4}} = 4$.

$$\text{b } \frac{4.7 \times 10^{27}}{3.9 \times 10^{26}} = \frac{R_{star}^2 (9250)^4}{R_{sun}^2 (6000)^4} \Rightarrow \frac{R_{star}}{R_{sun}} = \sqrt{\frac{4.7 \times 10^{27}}{3.9 \times 10^{26}} \times \frac{(6000)^4}{(9250)^4}} \approx 1.5$$

15 a Since $L = \sigma AT^4 = \sigma 4\pi R^2 T^4$:

$$\frac{5.2 \times 10^{28}}{3.9 \times 10^{26}} = \frac{R_{star}^2 (4000)^4}{R_{sun}^2 (6000)^4} \Rightarrow \frac{R_{star}}{R_{sun}} = \sqrt{\frac{5.2 \times 10^{28}}{3.9 \times 10^{26}} \times \frac{(6000)^4}{(4000)^4}} \approx 26$$

$$\text{b } \frac{b_A}{b_B} = \frac{\left(\frac{L_A}{4\pi d_A^2}\right)}{\left(\frac{L_B}{4\pi d_B^2}\right)} \text{ and so } 2 = \frac{d_B^2}{d_A^2} \Rightarrow \frac{d_A}{d_B} = 0.71.$$

16 We have that $b = \frac{L}{4\pi d^2}$ and $L = \sigma AT^4$. Combining, $b = \frac{\sigma AT^4}{4\pi d^2}$.

$$\text{Hence, } \frac{b_A}{b_B} = \frac{\left(\frac{\sigma AT_A^4}{4\pi d_A^2}\right)}{\left(\frac{\sigma AT_B^4}{4\pi d_B^2}\right)} = \frac{\left(\frac{T_A^4}{d_A^2}\right)}{\left(\frac{T_B^4}{d_B^2}\right)} = \frac{T_A^4 d_B^2}{T_B^4 d_A^2} \Rightarrow \frac{T_A}{T_B} = \sqrt[4]{\frac{b_A d_A^2}{b_B d_B^2}}. \text{ Since this is a ratio we do not have to change units}$$

$$\text{(light years to meters.) Hence, } \frac{T_A}{T_B} = \sqrt[4]{\frac{8.0 \times 10^{-13} 120^2}{2.0 \times 10^{-15} 150^2}} = 4.$$

17 The distance to the star is $\frac{1}{0.034} = 29.4$ pc = $29.4 \times 3.09 \times 10^{16} = 9.08 \times 10^{17}$ m. The apparent brightness is

$$\text{then } b = \frac{L}{4\pi d^2} = \frac{2.45 \times 10^{28}}{4\pi \times (9.08 \times 10^{17})^2} = 2.4 \times 10^{-9} \text{ W m}^{-2}.$$

D2 Stellar characteristics and stellar evolution

- 18 Light emitted from the star will have to pass through the outer layers of the star. Atoms in these layers may absorb light of certain wavelengths if these wavelengths correspond to energy differences in the atomic energy levels. The absorbed photons will therefore not make it through the outer layers of the star and will appear as dark lines in the spectrum of the star.
- 19 The dark lines in the absorption spectrum of a star indicate that photons of a wavelength corresponding to the dark lines have been absorbed by atoms in the outer layers of the star. Different atoms absorb different wavelength photons and so the dark lines are indicative of the type of elements present in the star.
- 20 The color of the star corresponds to a particular wavelength. This is the peak wavelength in the spectrum which in turn is related to surface temperature through Wien's law, $\lambda T = 2.9 \times 10^{-3}$ K m.

21 We know that $L = \sigma AT^4$ and so $\frac{L_A}{L_B} = \frac{\sigma A_A T_A^4}{\sigma A_B T_B^4}$. Since the radius of A is double that of B, $\frac{L_A}{L_B} = 4 \frac{T_A^4}{T_B^4}$. From

Wien's law, $\lambda T = \text{const}$ and so $650 \times T_A = 480 \times T_B \Rightarrow \frac{T_A}{T_B} = \frac{480}{650}$. Hence, $\frac{L_A}{L_B} = 4 \times \left(\frac{480}{650}\right)^4 = 1.2$.

- 22 a An HR diagram is a plot of the luminosity of star versus its surface temperature. Temperature is plotted increasing to the left on the horizontal axis.
- b Such a plot reveals that stars are grouped into large classes: the main sequence which occupies a diagonal strip from top right to bottom left; the white dwarfs in the lower left corner and the red giants and supergiants in the top right corner.
- c i The luminosity is $L = 10^4 L_\odot = 3.9 \times 10^{30}$ W and the radius is $R = 10 R_\odot = 7.0 \times 10^9$ m. Therefore

$$3.9 \times 10^{30} = 5.67 \times 10^{-8} \times 4\pi(7.0 \times 10^9)^2 \times T^4$$

$$T = \left(\frac{3.9 \times 10^{30}}{5.67 \times 10^{-8} \times 4\pi(7.0 \times 10^9)^2} \right)^{\frac{1}{4}}$$

$$T \approx 18000 \text{ K}$$

ii From the mass luminosity relation $\frac{L}{L_\odot} = 10^4 = \left(\frac{M}{M_\odot}\right)^{3.5} \Rightarrow \frac{M}{M_\odot} = 10^{\frac{4}{3.5}} = 13.9$. Hence the density is

$$\rho = \frac{M}{V} = \frac{13.9 M_\odot}{10^3 V_\odot} = 1.4 \times 10^{-2} \rho_\odot$$

- d Star B: the luminosity is the same as star A and the temperature is 3000 K. Hence

$$\frac{L_A}{L_B} = 1 = \frac{\sigma 4\pi R_A^2 T_A^4}{\sigma 4\pi R_B^2 T_B^4} = \frac{R_A^2 18000^4}{R_B^2 3000^4} \Rightarrow \frac{R_A}{R_B} \approx 2.8 \times 10^{-2}. \text{ So } R_B = \frac{R_A}{2.8 \times 10^{-2}} = \frac{10 R_\odot}{2.8 \times 10^{-2}} = 360 R_\odot.$$

Star C: the luminosity is $L = 10^{-3} L_\odot$ and the temperature is the same as that of star A. Hence

$$\frac{L_A}{L_C} = \frac{10^4 L_\odot}{10^{-3} L_\odot} = 10^7 = \frac{\sigma 4\pi R_A^2 T^4}{\sigma 4\pi R_C^2 T^4} = \frac{R_A^2}{R_C^2} \Rightarrow \frac{R_A}{R_C} = \sqrt{10^7} = 3.2 \times 10^3$$

$$R_C = \frac{R_A}{3.2 \times 10^3} = \frac{10 R_\odot}{3.2 \times 10^3} = 3.2 \times 10^{-3} R_\odot$$

- 23 The temperature of the star is found from Wien's law to be $T = \frac{2.9 \times 10^{-3}}{2.42 \times 10^{-7}} = 1.20 \times 10^4$ K. From the HR diagram this corresponds to a luminosity of about 20 solar luminosities. Therefore

$$d = \sqrt{\frac{20 \times 3.9 \times 10^{26}}{4\pi \times 8.56 \times 10^{-12}}} = 8.5 \times 10^{18} \text{ m.}$$

24 From the mass luminosity relation $L \propto M^{3.5}$ it follows that

$$\frac{L}{L_{\odot}} = \left(\frac{M}{M_{\odot}} \right)^{3.5} = 15^{3.5} = 1.3 \times 10^4.$$

25 a From the mass luminosity relation $L \propto M^{3.5}$ it follows that $\frac{L}{L_{\odot}} = 4500 = \left(\frac{M}{M_{\odot}} \right)^{3.5} \Rightarrow \frac{M}{M_{\odot}} = 4500^{\frac{1}{3.5}} \approx 11$.

b If it was its luminosity should have been $\frac{L}{L_{\odot}} = \left(\frac{M}{M_{\odot}} \right)^{3.5} = 12^{3.5} \approx 6000$. The actual luminosity is 3200 times that of the sun so this star cannot be a main sequence star.

27 The luminosity is about 4000 solar luminosities and so

$$d = \sqrt{\frac{4000 \times 3.9 \times 10^{26}}{4\pi \times 3.45 \times 10^{-14}}} = 1.9 \times 10^{21} \text{ m}$$

28 a The peak wavelength is about $\lambda = 0.40 \times 10^{-6} \text{ m}$ and so the surface temperature (from Wien's law) is

$$T = \frac{2.9 \times 10^{-3}}{0.40 \times 10^{-6}} \approx 7.2 \times 10^3 \text{ K.}$$

b From the H-R diagram the luminosity is about 5–8 times that of the sun.

29 The temperature (from Wien's law) is $T = \frac{2.9 \times 10^{-3}}{7 \times 10^{-7}} \approx 4 \times 10^3 \text{ K.}$

30 a The speed is given by $v = \frac{2\pi R}{T} = 2\pi Rf = 2\pi \times 30 \times 10^3 \times 500 = 3.1 \times 10^7 \text{ m s}^{-1}$.

b $\frac{3.1 \times 10^7}{3.0 \times 10^8} \approx 10\%$

31 A red giant forms out of a main sequence star when a certain percentage of the hydrogen of the star is used up in nuclear fusion reactions. The core of the star collapses and this releases gravitational potential energy that warms the core to sufficiently high temperatures for fusion of helium in the core to begin. The suddenly released energy forces the outer layers of the star to expand rapidly and to cool down. The star thus becomes a bigger but cooler star – a red giant.

32 a A planetary nebula refers to the explosion of a red giant star that ejects most of the mass of the star into space.

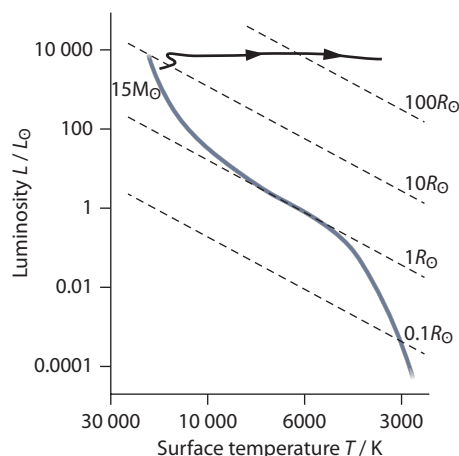
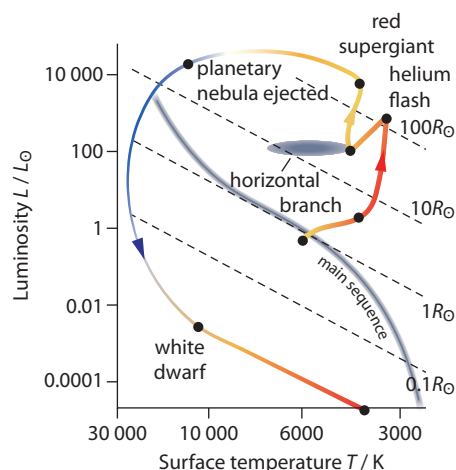
b Not all planetary nebulas (of the 3000 or so that are known to exist) appear as rings they way the famous helix and ring nebulas appear. The ring – like appearance is because the gas surrounding the center is very thin. A line of sight through the outer edges of the nebula goes through much more gas than a line of sight through the center. Hence the center looks transparent while the edges do not.

33 No, because all the elements that are necessary for life were made either in nuclear fusion processes in the cores of very heavy stars or during the supernova stage when nuclei were irradiated with neutrons (to make the elements heavier than iron).

34 a A 2 solar mass star would evolve to become a red giant. As the star expands in size into the red giant stage, nuclear reactions inside the core of the star are able to produce heavier elements than helium because the temperature of the core is sufficiently high. The red giant star will explode as a planetary nebula ejecting most of the mass of the star into space and leaving behind a dense core. The core is no longer capable of nuclear reactions and the star continues to cool down. The core is stable under further collapse because of electron degeneracy pressure. The core has a mass that is less than the Chandrasekhar limit and so ends up as a white dwarf.

b A 20 solar mass star would evolve to become a red supergiant. The red supergiant star will explode as a supernova ejecting most of the mass of the star into space and leaving behind a dense core. The core is no longer capable of nuclear reactions and the star continues to cool down. The core is stable under further collapse because of neutron degeneracy pressure. The core must have a mass that is less than Oppenheimer-Volkoff limit and so ends up as a neutron star.

c See graphs.



- 35 **a** A white dwarf forms as the core left behind after planetary nebula explosion of a red giant star.
b White dwarfs are very small (Earth size in radius) and very dense.
c A quantum mechanical principle known as the Pauli exclusion principle forbids neutrons to occupy the same quantum state. The enormous densities in neutrons stars try to force neutrons into the same state. A pressure develops among the neutrons to keep them apart and this balances the gravitational pressure.
- 36 Two from:
 1. A main sequence star provides energy by nuclear fusion; no fusion takes place in a white dwarf.
 2. With the exception of a few of the smallest main sequence stars (the red dwarfs) main sequence stars are larger in radius than a white dwarf.
 3. The density of main sequence stars is much less than that of white dwarfs.
 4. Main sequence stars are in equilibrium under the action of gravitational and radiation pressures whereas white dwarfs between gravitational and electron degeneracy pressures.
- 37 The density will be $\rho = \frac{1.0 \times 10^{30}}{\frac{4\pi}{3}(6.4 \times 10^6)^3} = 9.1 \times 10^8 \text{ kg m}^{-3}$.
- 38 Two from:
 1. A main sequence star provides energy by nuclear fusion; no fusion takes place in a neutron star.
 2. Even the smallest neutron star is larger in radius than a neutron star.
 3. The density of main sequence stars is much less than that of neutron stars.
 4. Main sequence stars are in equilibrium under the action of gravitational and radiation pressures whereas neutron stars between gravitational and neutron degeneracy pressures.
- 39 **a** A neutron star forms as the core left behind after a supernova in a red supergiant star.
b Neutron stars are very small (tens of km in radius) and very dense.
c A quantum mechanical principle known as the Pauli exclusion principle forbids neutrons to occupy the same quantum state. The enormous densities in neutrons stars try to force neutrons into the same state. A pressure develops among the neutrons to keep them apart and this balances the gravitational pressure.
- 40 This is mass of $1.4M_{\odot}$ and is the largest mass a white dwarf can have. A core with a mass larger than this limit cannot withstand the pressure of gravity by electron degeneracy pressure and will collapse further.
- 41 This is mass of $2 - 3M_{\odot}$ and is the largest mass a neutron star can have. A core with a mass larger than this limit cannot withstand the pressure of gravity by neutron degeneracy pressure and will collapse further, presumably without limit into a black hole.

(In 2011 the heaviest known neutron star was discovered using the National Radio Astronomy Observatory at Green Bank in Virginia in the US. The mass is $(1.97 \pm 0.04)M_{\odot}$. The neutron star is a member of the binary pulsar J1614–2230. The significance of this discovery is that it rules out exotic forms of matter that have been proposed to exist inside neutron stars.)

- 42 The gas law states that $PV = nRT$. The number of moles is equal to $n = \frac{N}{N_A}$ where N_A is Avogadro's number. The Boltzmann constant k is defined by $k = \frac{R}{N_A}$ and so the gas law may be written as $PV = NkT$.
- Since $V \propto R^3$ and $N \propto M$ we have that $PR^3 \propto MT$. From dimensional analysis, equilibrium demands that $PA \approx \frac{GM^2}{R^2}$ where A is the area on which the pressure P acts. Since $A \propto R^2$ it follows that $P \propto \frac{M^2}{R^4}$ and combining these two equations we get $T \propto \frac{M}{R}$. This shows that as the star shrinks (the radius gets smaller) the temperature goes up. Now, the luminosity is given by $L = \sigma AT^4 \propto R^2 \left(\frac{M}{R}\right)^4 = \frac{M^4}{R^2}$. And since $\rho \propto \frac{M}{R^3}$ i.e. $R \propto \left(\frac{M}{\rho}\right)^{1/3}$ it follows that $L \propto \frac{M^4}{M^{2/3}} = M^{3.3}$, the mass-luminosity relation!

D3 Cosmology

- 43 a The distant galaxies move away from earth with a speed that is proportional to their distance from earth.
 b This is evidence for the expanding universe because it implies that space in between galaxies is stretching meaning that the volume of the universe is increasing i.e. it expands.
- 44 No. These are nearby galaxies which show a blueshift because the mutual gravitational attraction between them and our Milky Way makes them move toward us.
- 45 Taking the Hubble constant to be $H = 68 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and Hubble's law, we find
- $$v = Hd \Rightarrow d = \frac{v}{H} = \frac{500}{68} \approx 7 \text{ Mpc.}$$
- 46 There is no empty previously unoccupied space into which the galaxies are moving. Space is being created in between the galaxies.
- 47 The big bang signifies the beginning of time and space. At the big bang the universe was a point and so the big bang happened everywhere in the universe.
- 48 The question is meaningless *within the big bang model* since **by definition** time started with the big bang. It is as meaningless as to ask for a place 1 km north of the north pole. However, recent developments within string theory suggest that the question may not be as meaningless as it appears. See the very interesting article "The time before time", by Gabriele Veneziano (one of the true greats of theoretical physics) in the May 2004 Scientific American.
- 49 No, because these are nearby galaxies whose motion is much more affected by their mutual gravitational attraction rather than by the cosmic expansion.
- 50 a $\frac{v}{c} = \frac{\Delta\lambda}{\lambda_0} \Rightarrow v = \frac{c\Delta\lambda}{\lambda_0} = \frac{3 \times 10^8 \times (658.9 - 656.3)}{656.3} \approx 1.2 \times 10^6 \text{ m s}^{-1}$.
- $v = Hd \Rightarrow d = \frac{v}{H} = \frac{1.2 \times 10^6}{68 \times 10^3} \approx 18 \text{ Mpc.}$ We have expressed the Hubble constant as $H = 72 \times 10^3 \text{ m s}^{-1} \text{ Mpc}^{-1}$ so that the distance comes out in Mpc.
- b $\frac{v}{c} = \frac{\Delta\lambda}{\lambda_0} \Rightarrow v = \frac{c\Delta\lambda}{\lambda_0} = \frac{3 \times 10^8 \times (689.1 - 656.3)}{656.3} \approx 1.5 \times 10^7 \text{ m s}^{-1}$.
- $v = Hd \Rightarrow d = \frac{v}{H} = \frac{1.5 \times 10^7}{68 \times 10^3} \approx 220 \text{ Mpc.}$
- c $\frac{v}{c} = \frac{\Delta\lambda}{\lambda_0} \Rightarrow v = \frac{c\Delta\lambda}{\lambda_0} = \frac{3 \times 10^8 \times (704.9 - 656.3)}{656.3} \approx 2.2 \times 10^7 \text{ m s}^{-1}$.
- $v = Hd \Rightarrow d = \frac{v}{H} = \frac{2.2 \times 10^7}{68 \times 10^3} \approx 320 \text{ Mpc.}$

$$d \quad \frac{v}{c} = \frac{\Delta\lambda}{\lambda_0} \Rightarrow v = \frac{c\Delta\lambda}{\lambda_0} = \frac{3 \times 10^8 \times (741.6 - 656.3)}{656.3} \approx 3.9 \times 10^7 \text{ m s}^{-1}.$$

$$v = Hd \Rightarrow d = \frac{v}{H} = \frac{3.9 \times 10^7}{68 \times 10^3} \approx 570 \text{ Mpc}.$$

$$e \quad \frac{v}{c} = \frac{\Delta\lambda}{\lambda_0} \Rightarrow v = \frac{c\Delta\lambda}{\lambda_0} = \frac{3 \times 10^8 \times (789.7 - 656.3)}{656.3} \approx 6.1 \times 10^7 \text{ m s}^{-1}.$$

$$v = Hd \Rightarrow d = \frac{v}{H} = \frac{6.1 \times 10^7}{68 \times 10^3} \approx 900 \text{ Mpc}.$$

$$51 \text{ a} \quad v = Hd \Rightarrow d = \frac{v}{H} = \frac{3.0 \times 10^8}{68 \times 10^3} \approx 4.4 \times 10^3 \text{ Mpc}.$$

b They are unobservable.

c No because the Hubble speed is not a real speed. It is the result of space in between galaxies stretching. It is not the speed of any material object.

52 The three standard pieces of evidence in favor of the big bang are (1) the cosmic background radiation, (2) the expansion of the universe and (3) the helium abundance in the universe.

$$53 \text{ a} \quad \text{The redshift is } \frac{\Delta\lambda}{\lambda_0} = \frac{5.3 \times 10^{-7} - 4.5 \times 10^{-7}}{4.5 \times 10^{-7}} = 0.178 \approx 0.18.$$

$$b \quad \frac{\Delta\lambda}{\lambda_0} = \frac{v}{c} \text{ so } v = 0.178c \approx 5.3 \times 10^4 \text{ km s}^{-1}.$$

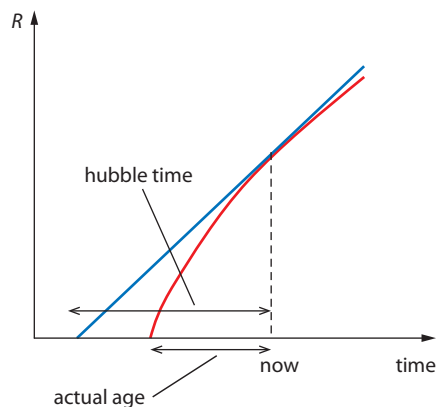
$$c \quad \text{From Hubble's law, } v = Hd, \text{ we have that } d = \frac{v}{H} = \frac{5.3 \times 10^4 \text{ km s}^{-1}}{68 \text{ km s}^{-1} \text{ Mpc}^{-1}} \approx 780 \text{ Mpc}.$$

54 At the time of the big bang the distance between any two objects was zero. A galaxy that today is at a distance d traveled this distance in time T , the age of the universe today and so $v = \frac{d}{T}$. The present speed of recession of the galaxy is $v = Hd$. Assuming that the galaxy had this speed throughout, then $\frac{d}{T} = Hd$, i.e. $T = \frac{1}{H}$.

$$\text{With } H = 500 \text{ km s}^{-1} \text{ Mpc}^{-1}, T = \frac{1}{500 \text{ km s}^{-1} \text{ Mpc}^{-1}} = \frac{1}{500 \times 10^3} 10^6 \times 3.26 \times 9.46 \times 10^{15} \text{ s, i.e.}$$

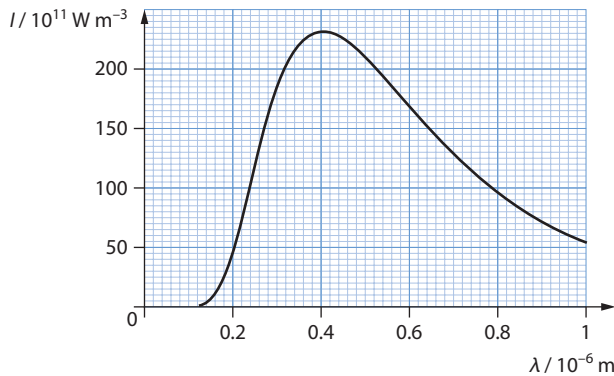
$T = 6.2 \times 10^{19} \text{ s}$ or 2 billion years. (The earth is older than this estimate.)

55 The estimate of the age of the universe as $T = \frac{1}{H}$ is based on the assumption that the galaxies have been moving away from earth at their present speed. The Hubble time is the age the Universe would have if it expanded at its present rate (blue line in the diagram below).

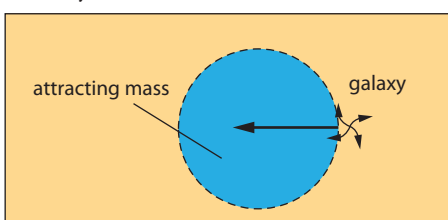


56 The fact that the speed of recession is proportional to the distance from earth implies that any other observer, anywhere else in the universe would reach the same conclusion, i.e. that he/she is at the center of the expansion. Thus, there is no center of expansion.

- 57 a The significance of the CMB is that it provides evidence for the Hot Big Bang theory. Radiation at a temperature of 2.7 k corresponds to a wavelength of about 1 mm. In a Hot Big Bang model radiation in the past would have been very high with a much smaller wavelength. As the Universe expands, space stretches and the wavelength of the radiation would increase as is observed.
- b The same.
- 58 a See graph.



- b $\lambda = \frac{2.9 \times 10^{-3}}{2.7} \approx 1.1 \times 10^{-3} \text{ m}$.
- 59 a It keeps decreasing approaching absolute zero.
- b It would reach a minimum at the largest size of the universe and then would begin to increase as the universe begins to collapse.
- 60 a Redshift is the ratio of the difference between the observed and emitted wavelengths to the emitted wavelength.
- b Space stretches in between galaxies and so does the wavelength,
- c $z = \frac{v}{c} = \frac{H_0 d}{c} \Rightarrow d = \frac{cz}{H_0}$.
- d $d = \frac{3.0 \times 10^5 \times 0.18}{68} \approx 790 \text{ Mpc}$.
- e $z = \frac{R}{R_0} - 1 \Rightarrow \frac{R}{R_0} = 1.18$.
- 61 a $d = \frac{cz}{H_0} = \frac{3.0 \times 10^5 \times \frac{15}{486}}{68} \approx 140 \text{ Mpc}$.
- b $z = \frac{R}{R_0} - 1 = \frac{15}{486} \Rightarrow \frac{R}{R_0} = 1.03$
- 62 $z = \frac{R}{R_0} - 1 = \frac{1}{0.85} = 0.176$. Hence $d = \frac{cz}{H_0} = \frac{3.0 \times 10^5 \times 0.176}{68} \approx 780 \text{ Mpc}$.
- 63 All type Ia supernovae have the same peak luminosity and so measuring its apparent brightness at the peak allows determining the distance.
- 64 a The speed of expansion of a distant galaxy is greater than what would have been assumed based on the simple Hubble law $v = Hd$.
- b Gravity should have slowed them down.



- c Distant Type Ia supernovae were used as standard candles i.e. objects of known luminosity. Measuring the apparent brightness at the peak allowed determination of distance. Measuring redshift allowed determination of the deceleration parameter of the universe because redshift and distance are related to each other through this parameter. High redshift values were needed and so very distant and very bright objects had to be used. Type Ia supernovae fitted these requirements.
- d To determine the distance the apparent brightness had to be measured at a time when the luminosity was at the peak because the peak the luminosity was known.
- 65 In a decelerating universe the distance to a distant supernova would have been smaller and so we would expect to see a brighter supernova.
- 66 By systematically scanning large areas of the sky over and over again and obtaining very large numbers of digital images that could be analysed by a computer. In this way very large numbers of galaxies were examined.

D4 Stellar processes

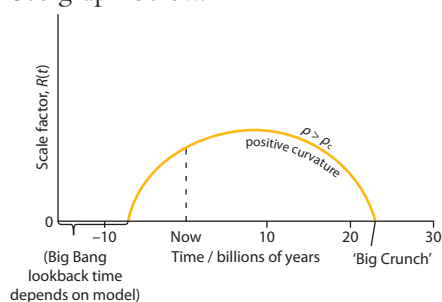
- 67 A cloud of dust will collapse and form a star if the gravitational potential energy of the cloud molecules is greater than their random kinetic energy.
- 68 Once the gravitational potential energy of the clouds molecules is higher than the kinetic energy the cloud will begin to contract. This will release gravitational potential energy that will heat the cloud to temperatures large enough for nuclear fusion to take place.
- 69 Cold; so that the kinetic energy of the molecules will be smaller than their gravitational potential energy.
- 70 The mass luminosity relation states that $L \propto M^{3.5}$. This means that massive stars have a disproportionately high luminosity. In other words: a star will leave the main sequence when it exhausts a certain fraction k of its hydrogen mass by nuclear fusion. Then, roughly, $L = \frac{kM}{T}$ and so $M^{3.5} \propto \frac{kM}{T} \Rightarrow T \propto M^{-2.5}$ meaning that higher mass spend less time T on the main sequence.
- 71 From the previous problem, $T \propto M^{-2.5}$. Hence $\frac{T}{T_{\odot}} = \left(\frac{M}{M_{\odot}}\right)^{-2.5} = \frac{1}{2^{2.5}} = \frac{1}{5.7}$.
- 72 The energy produced in the sun's lifetime is $E = 3.9 \times 10^{26} \times 10^{10} \times 365 \times 24 \times 3600 = 1.2 \times 10^{44}$ J. This corresponds to a mass of $m = \frac{E}{c^2} = \frac{1.2 \times 10^{44}}{9.0 \times 10^{16}} = 1.3 \times 10^{27}$ kg.
- 73 Because it marks the end of its life on the main sequence.
- 74 Once off the main sequence different sequences of nuclear fusion reactions take place depending on the mass of the star. The more massive the star the heavier the elements produced. The elements arrange themselves according to mass the heaviest being at the core and the lightest in the outer layers creating the onion like structure.
- 75 a On the main sequence the main nuclear reaction for low mass stars is the proton-proton cycle that produces helium by fusing hydrogen.
b A low mass star that leaves the main sequence will produce carbon in the triple alpha process in which three helium-4 nuclei produce carbon-12 (with the intermediate and subsequent fusing of beryllium).
- 76 a Such a massive star will produce helium both with the proton-proton cycle and the CNO cycle.
b After leaving the main sequence much heavier elements will be produced beginning with carbon in the triple alpha process and then neon, sodium, magnesium and silicon.
- 77 To produce elements by nuclear fusion requires the binding energy of the product nuclei to be higher than that of the reactants. Since iron is near the peak of the binding energy curve nothing heavier than iron can be produced by fusion.
- 78 a helium
b helium
c carbon

- 79 The s and r processes refer to neutron absorption by nuclei. Neutron absorption increases the mass number of a nucleus and the resulting isotope will in general decay by beta decay increasing the atomic number by 1 and thus producing a heavier element. In the slow s process, the isotope decays before the nucleus has time to absorb another neutron. In the fast r process the nucleus absorbs more than one neutron before it decays. The subsequent beta decay of the heavy isotope again produces heavier elements. The r process happens during a supernova explosion.
- 80 Heavier elements have higher atomic numbers and therefore higher positive electric charges. To fuse means that the nuclei have to come very close to each other and this requires higher kinetic energies. This in turn implies higher temperatures.
- 81 A Type Ia supernova is formed when a white dwarf that is in binary star system accretes mass from the companion star. The additional mass forces the white dwarf to exceed the Chandrasekhar limit and the star can no longer maintain equilibrium. The white dwarf contains mainly carbon and oxygen. The additional mass causes temperatures in the core to increase sufficiently for nuclear reactions to take place. The result is that the star blows up with enormous energy release.
- 82 A type II supernova is formed when a massive main sequence star that has evolved away from the main sequence and entered the red supergiant region explodes as supernova.
- 83 The main difference is their mechanism of production as described in the previous two problems. Additional differences include the shape of the light curve (i.e. the graph of luminosity versus time): the light curve for Type Ia falls off more sharply than that for Type II. Type II have hydrogen absorption lines whereas Type Ia do not.
- 84 The nuclear reactions inside massive stars produce an onion like structure with the heaviest elements at the core and the lightest, i.e. hydrogen, in the outer layers. Hence when the star goes supernova there is enough hydrogen present to produce hydrogen absorption lines.
- 85 Both have the net effect of turning 4 hydrogen nuclei into one nucleus of helium and both take place in main sequence stars. The CNO cycle requires higher temperatures and therefore takes place in more massive stars.

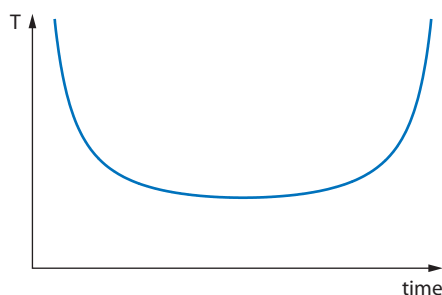
D5 Further cosmology

- 86 The cosmological principle states that the universe, on a large scale, is homogeneous (no one position is special) and isotropic (no one direction is special). This means that when viewed from different positions and different directions the observer sees the same distribution of matter and energy. **a** It can be used to deduce that the Universe has no centre for if it did observing from the centre would reveal a different picture than from any other point. **b** It can also be used to deduce that the Universe has no edge because, again, if it did viewing things from the edge would give a different picture than from a point far from the edge.
- 87 The main evidence for the cosmological principle is the isotropy of the CMB and the same distribution of galaxies in any direction from Earth. The fluctuations observed in the CMB are very small and so provide strong evidence for isotropy. Evidence for isotropy is also provided by the identical distribution of radio galaxies in different directions. Homogeneity is more difficult to test because we have not been able to move far from Earth so we have evidence only based in the small regions of the Universe that is our immediate neighbourhood.

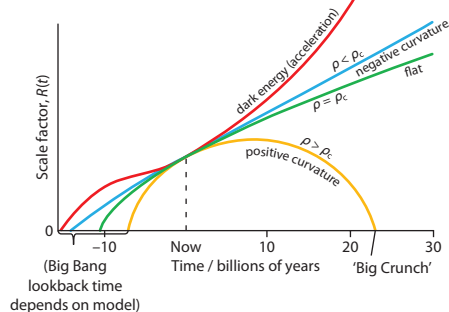
- 88 a A factor that relates the physical distance between two points in space in terms and the coordinates of the points.
 b See graph below.



- c Since the temperature depends on the scale factor according to $T \propto \frac{1}{R}$ we have a graph like this.



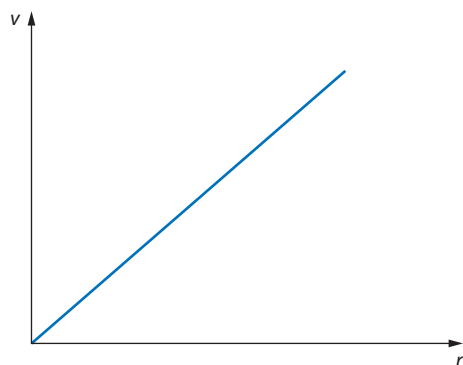
89



The graphs are arranged so that at the present time all models agree on the value of the Hubble constant because they have the same tangent line. The lines start at different times implying a different age of the universe depending on the model chosen.

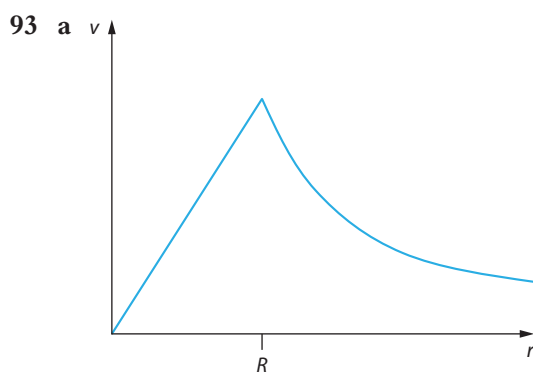
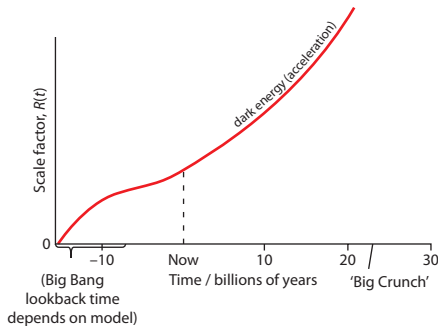
- 90 a Consider a cloud of uniform density ρ and a point particle moving in a circular orbit of radius r about the centre of the cloud. The speed of the particle is given by $v = \sqrt{\frac{GM}{r}}$ where M is the mass of the cloud within a sphere of radius r . We have that $M = \rho V = \rho \frac{4\pi r^3}{3}$ and so $v = \sqrt{\frac{G\rho \frac{4\pi r^3}{3}}{r}} = \sqrt{G\rho \frac{4\pi r^2}{3}}$ i.e. $v \propto r$.

- b See graph here.



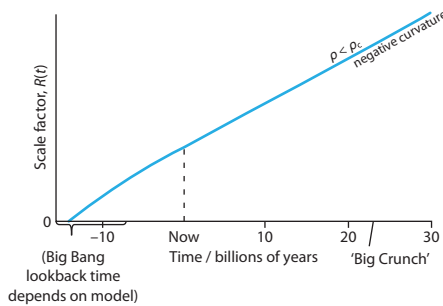
- 91 a The speed of the particle is given by $v = \sqrt{\frac{GM}{r}}$ and so $v = \sqrt{\frac{Gkr}{r}} = \sqrt{Gk}$ i.e. $v = \text{constant}$.
- b Far from the center the rotation speed approaches a constant consistent with a distribution $M(r) = kr$. This implies that there is substantial mass at large values of r i.e. in the outer edges of the galaxy.

92 See graph here.

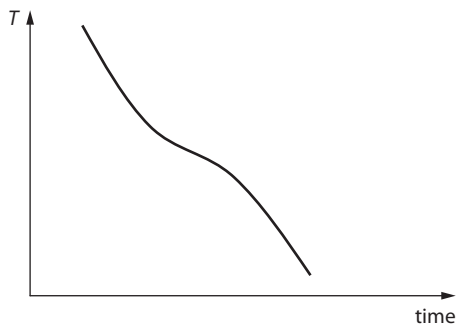


- b In the graph above the curve approaches zero while in our galaxy it approaches a non zero constant.
- 94 a Dark matter is matter that is known to exist because of its gravitational effects on the motion of other nearby matter but is too cold to radiate and so cannot be seen.
- b It could be baryonic matter i.e. matter consisting of ordinary protons and neutrons such as white, brown and black dwarfs as well as very small planets. However, baryonic matter by itself cannot account for all the dark matter that is known to exist. Other candidates include neutrinos and exotic particles predicted by supersymmetry.
- 95 Dark matter is matter that does not radiate and so cannot be seen. Dark energy is a vacuum energy that fills all space and is supposed to be responsible for a repulsive force that is accelerating the rate of expansion of the universe.
- 96 In an accelerating universe distant supernovae are further out than imagined and hence dimmer than what would be expected based on models in which the expansion slows down.
- 97 a The CMB radiation is isotropic to a very high degree which means that no matter in which direction one looks the spectrum of the CMB radiation is the same. However there are very small deviations from this perfect isotropy of the order of millionths of a kelvin: the temperature of the radiation is not constant but varies by these tiny amounts.
- b The anisotropies in the CMB are important because they are needed in order to explain the formation of structures in the universe such as stars and galaxies. In addition studies of the anisotropy also give information about various cosmological parameters most importantly about the curvature of the universe.
- 98 The magnitude of the fluctuations in temperature fluctuations is a direct function of the geometry of the universe. The measured values are consistent with a flat universe.
- 99 According to D44 every model with a negative cosmological constant (i.e. $\Omega_\Lambda < 0$) involve a collapsing universe which is not the case.

- 100 By the Wien displacement law $\lambda T = \text{constant}$. The cosmological origin of redshift implies that $\lambda \propto R$ and so $T \propto \frac{1}{R}$.
- 101 **a** In a model with zero dark energy (i.e. cosmological constant Λ) this is the density that separates a universe that expands forever to a universe that will eventually collapse. A universe with $\Lambda = 0$ and a density equal to the critical density would expand forever but with a rate that approaches zero at infinity.
- b** The total energy of a mass m a distance r from the centre of the cloud is $E = \frac{1}{2}mv^2 - \frac{GMm}{r}$ where M is the mass of the cloud. If we call the density of this cloud ρ , then $M = \rho \frac{4}{3}\pi r^3$ and using this together with $v = Hr$ we find $E = \frac{1}{2}mr^2 \left(H_0^2 - \frac{8\pi\rho G}{3} \right)$. The particle will move to infinity and just about stop if $E = 0$ i.e. at the critical density $\rho_c = \frac{3H^2}{8\pi G}$.
- c** The critical density is $\rho_c = \frac{3H^2}{8\pi G} = \frac{3 \times \left(\frac{68 \times 10^3}{10^6 \times 3.09 \times 10^{16}} \right)^2}{8\pi \times 6.67 \times 10^{-11}} = 8.7 \times 10^{-27} \approx 10^{-26} \text{ kg m}^{-3}$. Hence the matter density of the universe is $\Omega_m = \frac{\rho}{\rho_c} \Rightarrow \rho = 0.32 \times 8.7 \times 10^{-27} = 3 \times 10^{-27} \text{ kg m}^{-3}$.
- 102 **a** It is difficult because it involves estimating the mass in large volumes of space far from the Earth and this implies large uncertainties. In addition there is a lot of matter in the universe that does not radiate and so cannot be seen.
- b** In 1 m^3 we have a mass of 10^{-26} kg and so $\frac{10^{-26}}{1.67 \times 10^{-27}} \approx 6$ atoms of hydrogen.
- 103 **a** A model in which the presence of dark energy overcomes the retarding gravitational force making *the rate of change* of the scale factor to change a positive rate. The term refers specifically to a model with a positive cosmological constant or dark energy and flat curvature.
- b** See graph.



- c** See graph.



104 The surface of a flat sheet of paper extending forever is an example of what is called an open universe, whereas the surface of a sphere (just the surface not the interior) is an example of a closed universe. The surface of a sphere is finite (it has finite area) but it has no boundary – you cannot walk on a sphere and come to a point where you see an edge. On the other hand, a finite flat piece of paper is an example of a finite space (finite area) that does have an edge, a boundary.

105 Use $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} = \Omega_\Lambda \rho_c = 0.68 \times 10^{-26} \text{ kg m}^{-3}$. Hence

$$\Lambda = \frac{0.68 \times 10^{-26} \times 8\pi G}{c^2}$$

$$\Lambda = \frac{0.68 \times 10^{-26} \times 8\pi \times 6.67 \times 10^{-11} \text{ (N kg}^{-2} \text{ m}^2\text{)} (\text{kg m}^{-3}\text{)}}{9.0 \times 10^{16} \text{ m}^2 \text{ s}^{-2}}$$

$$\Lambda = 1.3 \times 10^{-52} \approx 10^{-52} \text{ m}^{-2}$$

106 a The universe is known to expand so R varies. The redshift of light from distant galaxies implies an increasing R . The existence of the CMB implies an increasing R .

b It could be anywhere on the blue line separating the expanding and collapsing phases in the D44 depending on the values chosen for the matter density and the cosmological constant.

107 We know that $\rho_c = \frac{3H^2}{8\pi G}$, $\Omega_m = \frac{\rho}{\rho_c}$ and $\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c}$ with $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$. Substituting in the

Friedmann equation $H^2 = \frac{8\pi G}{3} \left(\rho + \frac{\Lambda c^2}{8\pi G} \right) - \frac{kc^2}{R^2}$ gives

$$H^2 = \frac{8\pi G}{3} (\Omega_m + \Omega_\Lambda) \rho_c - \frac{kc^2}{R^2}$$

$$H^2 = \frac{H^2}{\rho_c} (\Omega_m + \Omega_\Lambda) \rho_c - \frac{kc^2}{R^2}$$

$$1 = (\Omega_m + \Omega_\Lambda) - \frac{kc^2}{H^2 R^2}$$

So if $\Omega_m + \Omega_\Lambda = 1$ it follows that $k = 0$, a flat universe.

APPENDIX A

Astronomical data

Body	Mass/kg	Radius/m	Orbit radius/m (average)	Orbital period
Sun	1.99×10^{30}	6.96×10^8	–	–
Moon	7.35×10^{22}	1.74×10^6	3.84×10^8	27.3 days
Mercury	3.30×10^{23}	2.44×10^6	5.79×10^{10}	88.0 days
Venus	4.87×10^{24}	6.05×10^6	1.08×10^{11}	224.7 days
Earth	5.98×10^{24}	6.38×10^6	1.50×10^{11}	365.3 days
Mars	6.42×10^{23}	3.40×10^6	2.28×10^{11}	687.0 days
Jupiter	1.90×10^{27}	6.91×10^7	7.78×10^{11}	11.86 yr
Saturn	5.69×10^{26}	6.03×10^7	1.43×10^{12}	29.42 yr
Uranus	8.66×10^{25}	2.56×10^7	2.88×10^{12}	83.75 yr
Neptune	1.03×10^{26}	2.48×10^7	4.50×10^{12}	163.7 yr
Pluto*	1.5×10^{22}	1.15×10^6	5.92×10^{12}	248.0 yr

Luminosity of the sun	$L = 3.9 \times 10^{26} \text{ W}$
Distance to nearest star (Proxima Centauri)	$4 \times 10^{16} \text{ m}$ (approx. 4.3 ly)
Diameter of the Milky Way	10^{21} m (approx. 100 000 ly)
Mass of the Milky Way	$4 \times 10^{41} \text{ kg}$
Distance to nearest galaxy (Andromeda)	$2 \times 10^{22} \text{ m}$ (approx. 2.3 million ly)

*Pluto has recently been downgraded into a new category of 'dwarf planet' (see Option D, Astrophysics).

APPENDIX B

Nobel prize winners in physics

No awards were made in years not listed.

2013: The prize was awarded jointly to François Englert and Peter W. Higgs “for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN’s Large Hadron Collider”

2012: The prize was awarded jointly to Serge Haroche and David J. Wineland “for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems”

2011: Half the prize was awarded to Saul Perlmutter, and the other half jointly to Brian P. Schmidt and Adam G. Riess “for the discovery of the accelerating expansion of the Universe through observations of distant supernovae”

2010: The prize was awarded jointly to Andre Geim and Konstantin Novoselov “for groundbreaking experiments regarding the two-dimensional material graphene”

2009: Half the prize was awarded to Charles Kuen Kao “for groundbreaking achievements concerning the transmission of light in fibers for optical communication” and the other half jointly to Willard S. Boyle and George E. Smith “for the invention of an imaging semiconductor circuit – the CCD sensor”

2008: Half the prize was awarded to Yoichiro Nambu “for the discovery of the mechanism of spontaneous broken symmetry in subatomic physics” and the other half jointly Makoto Kobayashi and Toshihide Maskawa “for the discovery of the origin of the broken symmetry which predicts the existence of at least three families of quarks in nature”

2007: The prize was awarded jointly to Albert Fert and Peter Grünberg “for the discovery of Giant Magnetoresistance”

2006: The prize was awarded jointly to John C. Mather and George F. Smoot (both USA) for their discovery of the black-body form and anisotropy of the cosmic microwave background radiation.

2005: Half the prize was awarded to Roy J. Glauber (USA) for his contribution to the quantum theory of optical coherence, and the other half was awarded jointly to John L. Hall (USA) and Theodor W. Hänsch (Germany) for their contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique.

2004: The prize was awarded jointly to D. J. Gross, H. D. Politzer and F. Wilczek (all USA) for their discovery of asymptotic freedom in quantum chromodynamics.



2003: The prize was awarded jointly to Alexei Abrikosov (Russia and USA), Vitaly Ginzburg (Russia) and Anthony Leggett (UK and USA) for pioneering contributions to the theory of superconductors and superfluids.

2002: Half the prize was awarded jointly to Raymond Davis Jr (USA) and Masatoshi Koshiba (Japan), and the other half was awarded to Riccardo Gianconi (USA) for pioneering contributions to astrophysics, particularly for the detection of cosmic neutrinos.

2001: The prize was awarded jointly to Eric Cornell (USA), Wolfgang Ketterle (Germany) and Carl Wieman (USA) for the achievement of Bose–Einstein condensation in dilute alkali gases and for early fundamental studies of the properties of the condensates.

2000: Half the prize was awarded jointly to Zhores I. Alferov (Russia) and Herbert Kroemer (USA) for developing semiconductor heterostructures used in high-speed and optoelectronics, and the other half was awarded to Jack St. Clair Kilby (USA) for his part in the invention of the integrated circuit.

1999: The prize was awarded jointly to Gerardus 't Hooft and Martinus J. G. Veltman (both Netherlands) for elucidating the quantum structure of electroweak interactions in physics.

1998: The prize was awarded jointly to Robert B. Laughlin (USA), Horst L. Stormer (Germany) and Daniel C. Tsui (USA) for their discovery of a new form of quantum fluid with fractionally charged excitations.

1997: The prize was awarded jointly to Steven Chu (USA), Claude Cohen-Tannoudji (France) and William D. Phillips (USA) for development of methods to cool and trap atoms with laser light.

1996: The prize was awarded jointly to David M. Lee, Douglas D. Osheroff and Robert C. Richardson (all USA) for their discovery of superfluidity in helium-3.

1995: The prize was awarded for pioneering experimental contributions to lepton physics, with half to Martin L. Perl (USA) for the discovery of the tau lepton, and the other half to Frederick Reines (USA) for the detection of the neutrino.

1994: The prize was awarded jointly to Bertram N. Brockhouse (Canada) and Clifford G. Shull (USA) for pioneering contributions to the development of neutron scattering techniques for studies of condensed matter: Brockhouse for the development of neutron spectroscopy, and Shull for the development of the neutron diffraction technique.

1993: The prize was awarded jointly to Russell A. Hulse and Joseph H. Taylor Jr (both USA) for the discovery of a new type of pulsar – a discovery that has opened up new possibilities for the study of gravitation.

1992: Georges Charpak (France) for his invention and development of particle detectors, in particular the multiwire proportional chamber.

1991: Pierre-Gilles de Gennes (France) for discovering that methods developed for studying order phenomena in simple systems can be generalized to more complex forms of matter, in particular to liquid crystals and polymers.

1990: The prize was awarded jointly to Jerome I. Friedman, Henry W. Kendall (both USA) and Richard E. Taylor (Canada) for their pioneering investigations concerning deep inelastic scattering of electrons on protons and bound neutrons, which have been of essential importance for the development of the quark model in particle physics.

1989: Half of the prize was awarded to Norman F. Ramsey (USA) for the invention of the separated oscillatory fields method and its use in the hydrogen maser and other atomic clocks, and the other half was awarded jointly to Hans G. Dehmelt (USA) and Wolfgang Paul (Germany) for the development of the ion trap technique.

1988: The prize was awarded jointly to Leon M. Lederman, Melvin Schwartz and Jack Steinberger (all USA) for the neutrino beam method and the demonstration of the doublet structure of the leptons through the discovery of the muon neutrino.

1987: The prize was awarded jointly to J. Georg Bednorz (Germany) and K. Alexander Müller (Switzerland) for their important breakthrough in the discovery of superconductivity in ceramic materials.

1986: Half of the prize was awarded to Ernst Ruska (Germany) for his fundamental work in electron optics and for the design of the first electron microscope, and the other half was awarded jointly to Gerd Binnig (Germany) and Heinrich Rohrer (Switzerland) for their design of the scanning tunnelling microscope.

1985: Klaus von Klitzing (Germany) for the discovery of the quantized Hall effect.

1984: The prize was awarded jointly to Carlo Rubbia (Italy) and Simon van der Meer (Netherlands) for their decisive contributions to the large project that led to the discovery of the field particles W and Z, communicators of the weak interaction.

1983: The prize was divided equally between Subrahmanyan Chandrasekhar (USA) for his theoretical studies of the physical processes of importance to the structure and evolution of the stars, and William A. Fowler (USA) for his theoretical and experimental studies of the nuclear reactions of importance in the formation of the chemical elements in the universe.

1982: Kenneth G. Wilson (USA) for his theory for critical phenomena in connection with phase transitions.

1981: Half the prize was awarded jointly to Nicolaas Bloembergen and Arthur L. Schawlow (both USA) for their contribution to the development of laser spectroscopy, and the other half was awarded to Kai M. Siegbahn (Sweden) for his contribution to the development of high-resolution electron spectroscopy.

1980: The prize was divided equally between James W. Cronin and Val L. Fitch (both USA) for the discovery of violations of fundamental symmetry principles in the decay of neutral K-mesons.



1979: The prize was divided equally between Sheldon L. Glashow (USA), Abdus Salam (Pakistan) and Steven Weinberg (USA) for their contributions to the theory of the unified weak and electromagnetic interaction between elementary particles, including, among other things, the prediction of the weak neutral current.

1978: Half the prize was awarded to Pyotr Leonidovich Kapitsa (USSR) for his basic inventions and discoveries in the area of low-temperature physics, and the other half was divided equally between Arno A. Penzias and Robert W. Wilson (both USA) for their discovery of cosmic microwave background radiation.

1977: The prize was divided equally between Philip W. Anderson (USA), Sir Nevill F. Mott (UK) and John H. van Vleck (USA) for their fundamental theoretical investigations of the electronic structure of magnetic and disordered systems.

1976: The prize was divided equally between Burton Richter and Samuel C. C. Ting (both USA) for their pioneering work in the discovery of a heavy elementary particle of a new kind.

1975: The prize was awarded jointly to Aage Bohr, Ben Mottelson (both Denmark) and James Rainwater (USA) for the discovery of the connection between collective motion and particle motion in atomic nuclei, and the development of the theory of the structure of the atomic nucleus based on this connection.

1974: The prize was awarded jointly to Sir Martin Ryle and Antony Hewish (both UK) for their pioneering research in radio astrophysics: Ryle for his observations and inventions, in particular of the aperture synthesis technique, and Hewish for his decisive role in the discovery of pulsars.

1973: Half the prize was equally shared between Leo Esaki (Japan) and Ivar Giaever (USA) for their experimental discoveries regarding tunnelling phenomena in semiconductors and superconductors, respectively, and the other half was awarded to Brian D. Josephson (UK) for his theoretical predictions of the properties of a supercurrent through a tunnel barrier, in particular those phenomena that are generally known as the Josephson effects.

1972: The prize was awarded jointly to John Bardeen, Leon N. Cooper and J. Robert Schrieffer (all USA) for their jointly developed theory of superconductivity, usually called the BCS theory.

1971: Dennis Gabor (UK) for his invention and development of the holographic method.

1970: The prize was divided equally between Hannes Alfvén (Sweden) for fundamental work and discoveries in magneto-hydrodynamics with fruitful applications in different parts of plasma physics, and Louis Néel (France) for fundamental work and discoveries concerning antiferromagnetism and ferromagnetism, which have led to important applications in solid state physics.

1969: Murray Gell-Mann (USA) for his contributions and discoveries concerning the classification of elementary particles and their interactions.

1968: Luis W. Alvarez (USA) for his decisive contributions to elementary particle physics, in particular the discovery of a large number of resonance states, made possible through his development of the technique of using a hydrogen bubble chamber and data analysis.

1967: Hans Albrecht Bethe (USA) for his contributions to the theory of nuclear reactions, especially his discoveries concerning energy production in stars.

1966: Alfred Kastler (France) for the discovery and development of optical methods for studying hertzian resonances in atoms.

1965: The prize was awarded jointly to Sin-Itiro Tomonaga (Japan), Julian Schwinger and Richard P. Feynman (both USA) for their fundamental work in quantum electrodynamics, with far-reaching consequences for the physics of elementary particles.

1964: Half the prize was awarded to Charles H. Townes (USA) and the other half was awarded jointly to Nicolay Gennadiyevich Basov and Aleksandr Mikhailovich Prokhorov (both USSR) for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser–laser principle.

1963: Half the prize was awarded to Eugene P. Wigner (USA) for his contributions to the theory of the atomic nucleus and the elementary particles, particularly through the discovery and application of fundamental symmetry principles, and the other half was awarded jointly to Maria Goeppert–Mayer (USA) and J. Hans D. Jensen (Germany) for their discoveries concerning nuclear shell structure.

1962: Lev Davidovich Landau (USSR) for his pioneering theories for condensed matter, especially liquid helium.

1961: The prize was divided equally between Robert Hofstadter (USA) for his pioneering studies of electron scattering in atomic nuclei and for his thereby achieved discoveries concerning the structure of the nucleons, and Rudolf Ludwig Mössbauer (Germany) for his researches concerning the resonance absorption of gamma radiation and his discovery in this connection of the effect which bears his name.

1960: Donald A. Glaser (USA) for the invention of the bubble chamber.

1959: The prize was awarded jointly to Emilio Gino Segre and Owen Chamberlain (both USA) for their discovery of the antiproton.

1958: The prize was awarded jointly to Pavel Alekseyevich Cherenkov, Il'ja Mikhailovich Frank and Igor Yevgenyevich Tamm (all USSR) for the discovery and the interpretation of the Cherenkov effect.

1957: The prize was awarded jointly to Chen Ning Yang and Tsung–Dao Lee (both China) for their penetrating investigation of the so-called parity laws, which has led to important discoveries regarding the elementary particles.

1956: The prize was awarded jointly, one-third each, to William Shockley, John Bardeen and Walter Houser Brattain (all USA) for their researches on semiconductors and their discovery of the transistor effect.



1955: The prize was divided equally between Willis Eugene Lamb (USA) for his discoveries concerning the fine structure of the hydrogen spectrum and Polykarp Kusch (USA) for his precision determination of the magnetic moment of the electron.

1954: The prize was divided equally between Max Born (UK) for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wavefunction, and Walther Bothe (Germany) for the coincidence method and his discoveries made using this method.

1953: Frits (Frederik) Zernike (Netherlands) for his demonstration of the phase contrast method, especially for his invention of the phase contrast microscope.

1952: The prize was awarded jointly to Felix Bloch and Edward Mills Purcell (both USA) for their development of new methods for nuclear magnetic precision measurements and discoveries made using these methods.

1951: The prize was awarded jointly to Sir John Douglas Cockcroft (UK) and Ernest Thomas Sinton Walton (Ireland) for their pioneering work on the transmutation of atomic nuclei by artificially accelerated atomic particles.

1950: Cecil Frank Powell (UK) for his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method.

1949: Hideki Yukawa (Japan) for his prediction of the existence of mesons on the basis of theoretical work on nuclear forces.

1948: Lord Patrick Maynard Stuart Blackett (UK) for his development of the Wilson cloud chamber method, and his discoveries using this method in the fields of nuclear physics and cosmic radiation.

1947: Sir Edward Victor Appleton (UK) for his investigations of the physics of the upper atmosphere, especially for the discovery of the so-called Appleton layer.

1946: Percy Williams Bridgman (USA) for the invention of an apparatus to produce extremely high pressures, and for the discoveries he made using this apparatus in the field of high-pressure physics.

1945: Wolfgang Pauli (Austria) for the discovery of the exclusion principle, also called the Pauli principle.

1944: Isidor Isaac Rabi (USA) for his resonance method for recording the magnetic properties of atomic nuclei.

1943: Otto Stern (USA) for his contribution to the development of the molecular ray method and his discovery of the magnetic moment of the proton.

1939: Ernest Orlando Lawrence (USA) for the invention and development of the cyclotron and for results obtained with it, especially with regard to artificial radioactive elements.

1938: Enrico Fermi (Italy) for his demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons.

1937: The prize was awarded jointly to Clinton Joseph Davisson (USA) and Sir George Paget Thomson (UK) for their experimental discovery of the diffraction of electrons by crystals.

1936: The prize was divided equally between Victor Franz Hess (Austria) for his discovery of cosmic radiation, and Carl David Anderson (USA) for his discovery of the positron.

1935: Sir James Chadwick (UK) for the discovery of the neutron.

1933: The prize was awarded jointly to Erwin Schrödinger (Austria) and Paul Adrien Maurice Dirac (UK) for the discovery of new productive forms of atomic theory.

1932: Werner Heisenberg (Germany) for the creation of quantum mechanics, the application of which has, among other things, led to the discovery of the allotropic forms of hydrogen.

1930: Sir Chandrasekhara Venkata Raman (India) for his work on the scattering of light and for the discovery of the effect named after him.

1929: Prince Louis-Victor de Broglie (France) for his discovery of the wave nature of electrons.

1928: Sir Owen Willans Richardson (UK) for his work on the thermionic phenomenon, and especially for the discovery of the law named after him.

1927: The prize was divided equally between Arthur H. Compton (USA) for his discovery of the effect named after him, and Charles Thomson Rees Wilson (USA) for his method of making the paths of electrically charged particles visible by condensation of vapour.

1926: Jean B. Perrin (France) for his work on the discontinuous structure of matter, and especially for his discovery of sedimentation equilibrium.

1925: The prize was awarded jointly to James Franck and Gustav Hertz (Germany) for their discovery of the laws governing the impact of an electron upon an atom.

1924: Karl Manne Georg Siegbahn (Sweden) for his discoveries and research in the field of X-ray spectroscopy.

1923: Robert Andrews Millikan (USA) for his work on the elementary charge of electricity and on the photoelectric effect.

1922: Niels Bohr (Denmark) for his services in the investigation of the structure of atoms and of the radiation emanating from them.

1921: Albert Einstein (Germany) for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect.

1920: Charles Edouard Guillaume (Switzerland) in recognition of the service he has rendered to precision measurements in physics by his discovery of anomalies in nickel–steel alloys.

1919: Johannes Stark (Germany) for his discovery of the Doppler effect in canal rays and the splitting of spectral lines in electric fields.

1918: Max Karl Ernst Ludwig Planck (Germany) in recognition of the services he rendered to the advancement of physics by his discovery of energy quanta.



1917: Charles Glover Barkla (UK) for his discovery of the characteristic Röntgen radiation of the elements.

1915: The prize was awarded jointly to Sir William Henry Bragg and Sir William Lawrence Bragg (both UK) for their services in the analysis of crystal structure by means of X-rays.

1914: Max von Laue (Germany) for his discovery of the diffraction of X-rays by crystals.

1913: Heike Kamerlingh-Onnes (Netherlands) for his investigations on the properties of matter at low temperatures, which led, among other things, to the production of liquid helium.

1912: Nils Gustaf Dalén (Sweden) for his invention of automatic regulators for use in conjunction with gas accumulators for illuminating lighthouses and buoys.

1911: Wilhelm Wien (Germany) for his discoveries regarding the laws governing the radiation of heat.

1910: Johannes Diderik van der Waals (Netherlands) for his work on the equation of state for gases and liquids.

1909: The prize was awarded jointly to Guglielmo Marconi (Italy) and Carl Ferdinand Braun (Germany) in recognition of their contributions to the development of wireless telegraphy.

1908: Gabriel Lippmann (France) for his method of reproducing colours photographically based on the phenomenon of interference.

1907: Albert Abraham Michelson (USA) for his optical precision instruments and the spectroscopic and metrological investigations carried out with their aid.

1906: Sir Joseph John Thomson (UK) in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases.

1905: Philipp Eduard Anton Lenard (Netherlands) for his work on cathode rays.

1904: Lord John William Strutt Rayleigh (UK) for his investigations of the densities of the most important gases and for his discovery of argon in connection with these studies.

1903: Half the prize was awarded to A. Henri Becquerel (France) in recognition of the extraordinary services he has rendered by his discovery of spontaneous radioactivity, and the other half was awarded jointly to Pierre and Marie Curie (France) in recognition of the extraordinary services they rendered by their joint researches on the radiation phenomena discovered by Henri Becquerel.

1902: The prize was awarded jointly to Hendrik A. Lorentz and Pieter Zeeman (both Netherlands) in recognition of the extraordinary service they rendered by their researches into the influence of magnetism upon radiation phenomena.

1901: Wilhelm K. Röntgen (Germany) in recognition of the extraordinary services he rendered by the discovery of the remarkable rays subsequently named after him.

Glossary

- A scan** a graph of reflected ultrasound signal strength versus time
- accelerating universe** recent measurements indicate that the rate of expansion of the universe is increasing rather than decreasing
- acoustic impedance** the product of the density of a substance and the speed of sound in that substance
- adiabatic** a change in which no heat enters or leaves a system
- age of the universe** the time from the Big Bang to the present time
- angular acceleration** the rate of change of angular velocity
- angular magnification** the ratio of the angle subtended by the image to the angle subtended by the object
- apparent brightness** the received power per unit area
- Archimedes' principle** the upthrust force on a body totally or partially immersed in a fluid is equal to the weight of the fluid that has been displaced by the body
- astronomical unit** the average radius of the Earth's orbit around the Sun
- attenuation** the loss of energy as radiation passes through matter
- attenuation coefficient** the probability per unit length that a particular photon will be absorbed
- B scan** a two-dimensional ultrasound image formed by putting together many A scans
- bending of light** when light bends in the curved space around a massive body
- Bernoulli effect** the lift on a wing when air above the wing flows faster than air under it
- Bernoulli equation** the equation relating the pressure, speed and height in the steady flow of a fluid
- Big Bang model** the prevailing model of the universe in which space, time, mass and energy were all created about 14 billion years ago
- binary star** a system of two stars orbiting a common centre
- black hole** a point in space where the curvature is infinite; an end product in the stellar evolution of very massive stars
- Carnot cycle** a cycle in a pressure–volume diagram consisting of two isothermal and two adiabatic changes
- Cepheid variables** stars with a periodic variation in the apparent brightness caused by expansions and contractions of the star's surface
- Chandrasekhar limit** the maximum mass of a white dwarf star, about 1.4 solar masses
- chromatic aberration** a defect of lenses due to the dependence of the refractive index on wavelength that leads to coloured images
- clock synchronisation** the process by which all clocks in a given frame are arranged to show the same time
- cluster of galaxies** a group of galaxies attracting each other gravitationally
- compound microscope** an instrument that magnifies images using two converging lenses
- conservation of angular momentum** when the net external torque on a system is zero, the total angular momentum of the system is constant
- constellation** a group of stars forming a recognisable pattern
- contrast** use of X-ray absorbing materials to line organs so that they can better be seen in an X-ray image
- converging lens** a lens where a set of incident rays parallel to the principal axis refract through the lens passing through a point on the principal axis on the other side of the lens
- concave mirror** a mirror in which a set of rays parallel to the principal axis reflect such that they pass through a point on the principal axis in front of the mirror
- cosmic background radiation** black body radiation in the microwave region at a temperature of about 2.7 K filling the Universe
- cosmological principle** the principle according to which the Universe, on a very large scale, is uniform and isotropic
- cosmological redshift** the interpretation of the observed red-shift in the light of distant galaxies in terms of space stretching in-between the source and the observer
- critical density** the density for which the rate of expansion of a flat (zero curvature) universe with zero cosmological constant approaches zero as time approaches infinity.
- curvature** the 'bending' of spacetime according to the mass and energy it contains so that spacetime does not obey the rules of Euclidean geometry
- damping** the presence of resistive forces that result in a decrease in the amplitude due to loss of energy in an oscillating system
- dark energy** energy that fills the universe and is thought to be responsible for the observed accelerated rate of expansion of the universe
- dark matter** matter that is too cold to radiate but is inferred to exist because of its gravitational effects
- decibel scale** a logarithmic scale on which intensity is measured
- density** the ratio of mass to volume
- diverging lens** a lens where a set of incident rays parallel to the principal axis refract through the lens such that their extensions pass through a point on the principal axis on the same side of the lens as the incident rays
- convex mirror** a mirror in which a set of rays parallel to the principal axis reflect such that their extensions pass through a point on the principal axis behind the mirror
- entropy** a measure of the disorder of a system
- equation of continuity** in laminar flow, the product of fluid speed and cross-sectional area of the tube is constant
- equivalence principle** states that it is impossible to distinguish effects of gravity from those of acceleration
- event horizon** a surface around a black hole where the escape speed is the speed of light
- first law of thermodynamics** the heat supplied to a system is equal to the change in the internal energy plus the work done
- flowtube** a set of neighbouring streamlines

- fluctuations in CMB** minute variations in the temperature of the cosmic background radiation (CMB)
- focal length** the distance from the centre of a lens or a mirror to the focal point
- focal point** for a lens, it is the point on the principal axis to which a set of refracted rays (or their extensions) converge when rays parallel to the principal axis are incident on the lens; for a mirror, it is the equivalent point for the set of reflected rays (or their extensions)
- galaxy** a very large number of stars bound together in one very large body
- Galilean transformation** the equations relating measurements in two reference frames moving past each other at constant velocity according to classical mechanics
- geodesic** a path of least length in curved space
- graded-index fibre** an optic fibre in which the refractive index of the core decreases gradually away from the core centre
- gradient field** additional magnetic field to which a patient is exposed in MRI in order to determine the point of absorption of the RF radiation
- gravitational red-shift** the decrease in the frequency of light as it rises in a gravitational field
- half-value thickness** the distance moved through a material at which the intensity is reduced by a factor of two
- Hertzsprung–Russell diagram** a plot of luminosity versus temperature for stars
- Hubble's law** distant galaxies move away from each other with speeds proportional to their separations
- hydrostatic equilibrium** the state of zero net force and zero net torque of a system immersed in a fluid
- ideal fluid** a theoretical fluid that is incompressible, has no viscosity and flows with laminar flow
- invariant** a quantity that has the same value in two different frames
- isobaric** a change in which the pressure is kept constant
- isothermal** a change in which the temperature is kept constant
- isovolumetric** a change in which the volume is kept constant
- Jean's criterion** a dust cloud will collapse and form a protostar when the gravitational potential energy of the particles making up the cloud is greater than their kinetic energy
- laminar flow** smooth flow with velocities of adjacent fluid layers parallel; the velocity at each point in the fluid is constant in time
- length contraction** the phenomenon in which a moving length is shorter when compared to a similar length at rest
- light year** the distance travelled by light in one year
- linear magnification** the ratio of image length to object length
- Lorentz transformation** the equations relating measurements in two reference frames moving past each other at constant velocity according to relativity
- luminosity** the total power radiated by a star
- magnetic resonance imaging (MRI)** imaging method using the phenomenon of nuclear magnetic resonance
- main sequence** a region of the Hertzsprung–Russell diagram from top left to bottom right containing stars undergoing fusion of hydrogen to helium
- mass absorption coefficient** the ratio of the linear attenuation coefficient to the density of the material
- mass–luminosity relation** the luminosity of main sequence stars is proportional to a power of their mass
- material dispersion** the dependence of the refractive index on wavelength in an optic fibre, which leads to different travel times for different wavelengths
- Minkowski diagram** another name for a spacetime diagram
- moment of inertia** a property of rigid, extended bodies that has to do with the distribution of mass around an axis
- muon decay** experiments in support of time dilation and length contraction
- natural frequency** the frequency of oscillation of an isolated system
- near point** the smallest distance at which the eye can focus without strain
- neutron capture** absorption of neutrons by nuclei
- neutron star** an end product in stellar evolution in which neutron degeneracy pressure is in equilibrium with gravitational pressure
- normal adjustment** for a telescope in normal adjustment, the final image is formed at infinity; for a microscope, the final image is formed at the near point
- nucleosynthesis** the processes by which the elements were produced
- Oppenheimer–Volkoff limit** the maximum mass of a neutron star, about three solar masses
- optic fibre** a thin tube in which light can propagate through successive total internal reflections
- parallax** a method to measure distances that uses the fact that an object looks shifted relative to a distant background when viewed from two different positions
- parsec** the distance at which the angle subtended by a length equal to one astronomical unit is one arc second
- Pascal's principle** a change in pressure applied to a point in an enclosed incompressible fluid is transmitted to all other parts of the fluid and its container
- planetary nebula** the ejection of mass from a red giant star
- postulates of relativity** in all inertial reference frames the speed of light in vacuum is the same; in all inertial reference the laws of physics are the same
- Pound–Rebka experiment** the experiment in which gravitational red-shift was first observed
- principal axis** an imaginary line passing through the centre of a lens or a mirror and normal to it
- proton spin relaxation** time taken for an excited proton to return to the ground state
- Q factor** a dimensionless number related to the amount of damping in a system that is equal to $2\pi \times \frac{\text{energy stored}}{\text{energy}}$ lost per cycle; the higher the Q factor, the longer the system oscillates before stopping
- R process** rapid absorption of neutrons by nuclei building elements heavier than bismuth-209
- radiation pressure** the outward pressure in a star created as a result of the energy produced in the star's core
- radio interferometry** the formation of an image using more than one radio telescope by combining the individual images



- radio telescope** a telescope that forms images by processing received radio waves
- ray diagram** a diagram showing the paths of refracted rays through a lens or the rays reflected off a mirror
- real image** an image formed by the intersection of actual rays
- red giant** very large, cool, reddish star with large luminosity
- reference frame** a coordinate system with clocks at every point in space
- reflecting telescope** a telescope that forms images using mirrors and reflection
- refracting telescope** a telescope that forms images using lenses and refraction
- relativistic momentum** the product of mass, velocity and the Lorentz gamma factor
- resonance** the state when the frequency of an externally applied periodic force equals the natural frequency of the system
- rest energy** the energy required to create a particle out of the vacuum
- rest frame** the frame of reference in which an object is at rest
- Reynold's number** dimensionless number characterising laminar (low values) or turbulent flow (high values)
- rotation curves** graphs of rotation speeds of galaxies versus radial distance
- rotational equilibrium** the condition that the net torque on a system is zero
- S process** slow absorption of neutrons by nuclei building elements up to bismuth-209
- scale factor** a term taken to roughly indicate the 'radius' of the universe
- second law of thermodynamics** the entropy of the universe always increases
- sharpness** the ability to see edges of different organs or different types of tissue
- simultaneity** events that are simultaneous in one frame and happen at different points in space will not be simultaneous in other frames
- spacetime diagram** a diagram of time and space coordinates used to show the position and time of events
- spectral class** a classification of stars based on their surface temperature and colour
- speed of light** the limiting speed that cannot be reached or exceeded by any material body
- spherical aberration** a defect of lenses and mirrors due to rays far from the principal axis having different focal lengths that leads to distortions in the image
- star** main sequence stars are spherical gaseous masses consisting mostly of hydrogen that are in equilibrium between gravitational pressure and radiation pressure; off the main sequence, stars have various compositions and means of equilibrium
- stellar cluster** a group of stars sufficiently close to each other to be attracting each other gravitationally
- stellar evolution** the processes by which a main sequence star leaves the main sequence and ends up in a final stage
- stellar spectra** sets of wavelengths that can be emitted by stars
- step-index fibre** an optic fibre in which the refractive index changes abruptly between core and cladding
- Stokes' law** the law giving the drag force on a spherical body moving through a fluid; the drag force is proportional to the speed and radius of the body
- streamlines** imaginary curves, tangents to which give the velocity vectors in fluid flow
- supercluster** a large number of clusters of galaxies
- thermal efficiency** the ratio of useful mechanical work done to the input energy
- time dilation** the phenomenon in which a moving clock runs slow when compared to a similar clock at rest
- torque** the product of force and the perpendicular distance between the line of action of the force and the rotation axis
- total energy** the sum of the rest and the kinetic energy of a particle
- translational equilibrium** the condition that the net force on a system is zero
- turbulence** the phenomenon of turbulent flow
- turbulent flow** fluid flow with velocities and densities varying wildly from point to point
- twin paradox** the 'paradox' where an astronaut leaving a twin behind on Earth returns after a long trip – according to the Earth-bound twin, the astronaut must be younger, but according to the astronaut, the Earth and the twin moved away and returned so the Earth-bound twin must be younger
- type Ia supernova** the increase in luminosity when material from one star in a binary star system falls into the other star (which is usually a white dwarf), initiating fusion
- type II supernova** the increase in luminosity when a massive red super giant star explodes
- ultrasound** sound of frequency higher than 20 kHz
- velocity addition** the formula relating speeds in two reference frames
- virtual image** an image formed by the intersection of extensions of rays
- viscosity** roughly, a measure of how resistive the fluid is to flowing motion
- waveguide dispersion** rays entering the fibre at different angles follow different paths and hence have different travel times
- white dwarf** an end product in stellar evolution in which electron degeneracy pressure is in equilibrium with gravitational pressure
- worldline** the path of a particle as shown on a spacetime diagram
- X-rays** electromagnetic radiation with a typical wavelength of 10^{-10} m

Acknowledgements

The authors and publishers acknowledge the following sources of copyright material and are grateful for the permissions granted. While every effort has been made, it has not always been possible to identify the sources of all the material used, or to trace all copyright holders. If any omissions are brought to our notice, we will be happy to include the appropriate acknowledgements on reprinting.

Artwork illustrations throughout © Cambridge University Press.

The chapter on Nature of Science was prepared by Dr Peter Hoeben.

The publisher would like to thank Neil Hodgson of Sha Tin College, Hong Kong, and Leigh Byrne of Cambridge House Grammar School, Ballymena, Northern Ireland for reviewing the content of this sixth edition.

Option A

p. 3 Image Asset Management/Alamy; p. 4 Peter Horree/Alamy; p. 8 Keystone Pictures USA/Alamy; p. 34 adapted from N. David Mermin, *It's About Time*, Princeton University Press, 2005; p. 51 The Print Collector/Alamy; p. 53_t Emilio Segre Visual Archives/American Institute of Physics/SPL; p. 53_b Mary Evans Picture Library/Alamy; p. 56_t, 56_m from M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004, © The Open University, used with permission; p. 63 European Southern Observatory/SPL.

Option B

p. 51 Library of Congress/SPL.

Option C

p. 13 EPA European Pressphoto Agency b.v./Alamy; p. 24 SPL; p. 25_t Prisma Bildagentur AG/Alamy; p. 25_b NASA 2002 http://hubblesite.org/gallery/spacecraft/05/web_print; p. 26 Image Asset Management Ltd/Alamy; p. 33_t Sergey Galushko/Alamy; p. 33_b GIPhotoStock/SPL.

Option D

p. 1_t Stocktrek Images, Inc/Alamy; p.1_m NASA, ESA, and the Hubble Heritage Team (STScI/AURA); p.1_{bl} NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration; p. 1_{br} NASA/ESA/STScI; p. 4 Keystone Pictures USA/Alamy; p. 13 Stocktrek Images, Inc./Alamy; p. 14 NASA, H.E. Bond and E. Nelan (Space Telescope Science Institute, Baltimore, Md.); M. Barstow and M. Burleigh (University of Leicester, U.K.); and J.B. Holberg (University of Arizona); p. 17 NASA, ESA, C.R. O'Dell (Vanderbilt University), and M. Meixner, P. McCullough; p. 18 Physics Today Collections/American Institute of Physics/SPL; p. 28_t adapted from graph 'Low redshift type 1a template lightcurve', Supernova Cosmology Project/Adam Reiss; p. 28_b NASA and A. Reiss (STScI); p. 35_l NASA, ESA, J. Hester and A. Loll (Arizona State University); p. 35_r NASA, ESA, and the Hubble Heritage Team (STScI/AURA) Acknowledgment: J. Hughes (Rutgers University); p. 38_t The Boomerang Collaboration; p. 38_m ESA and the Planck Collaboration; p. 38_{bl}, 42_t, 42_m from M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004, © The Open University, used with permission; p. 39_b (Adapted from Combes, F. (1991) Distribution of CO in the Milky Way, *Annual Review of Astronomy and Astrophysics*, 29, pp.195–237).

Thanks to The Open University for permission to reproduce graphs and illustrations from M. Jones and R. Lambourne, *An Introduction to Galaxies and Cosmology*, Cambridge University Press, in association with The Open University, 2004, © The Open University.

Key

l = left, *r* = right, *t* = top, *b* = bottom, *c* = centre
SPL = Science Photo Library