PHYSICS IN EVERYDAY LIFE

THE WORLD OF SCIENCE

bteiM I'cttt 1'utlaJo
DrMfureAvtli Kiiuilci,
tragic MwMilliii, Chili Mualir.
Nilil>>vr>

BcHiic Edilot.Mii>' >'**l*
IVWn CooviluM Jiitm Kdpewry
Prwctl Dirrcloc L»"irnl« Curtr

CantnbMir« Editor
Chr.xiiif'Suttnn

Adiivjt"
5it AUn G;IvitU IKS- MMKI of
Join 0:lkpr. Qul*™%c
Stirai Waning, UmttiBy of
Aumo.Tco*

Tamil-won
IfcrulItiiVfniinCM*;
DiM.G. H..t.la(7.8>

UiUI«cy<6)
J*kMi*fO«i(t)
aBi^KSunontll.lAU.H.lS.
16)

(iinpiirci gitcmcJ mokculic

Ownrd, UDn£Ob(m*X «diMiin'

4 AN EQUINOX BOOK

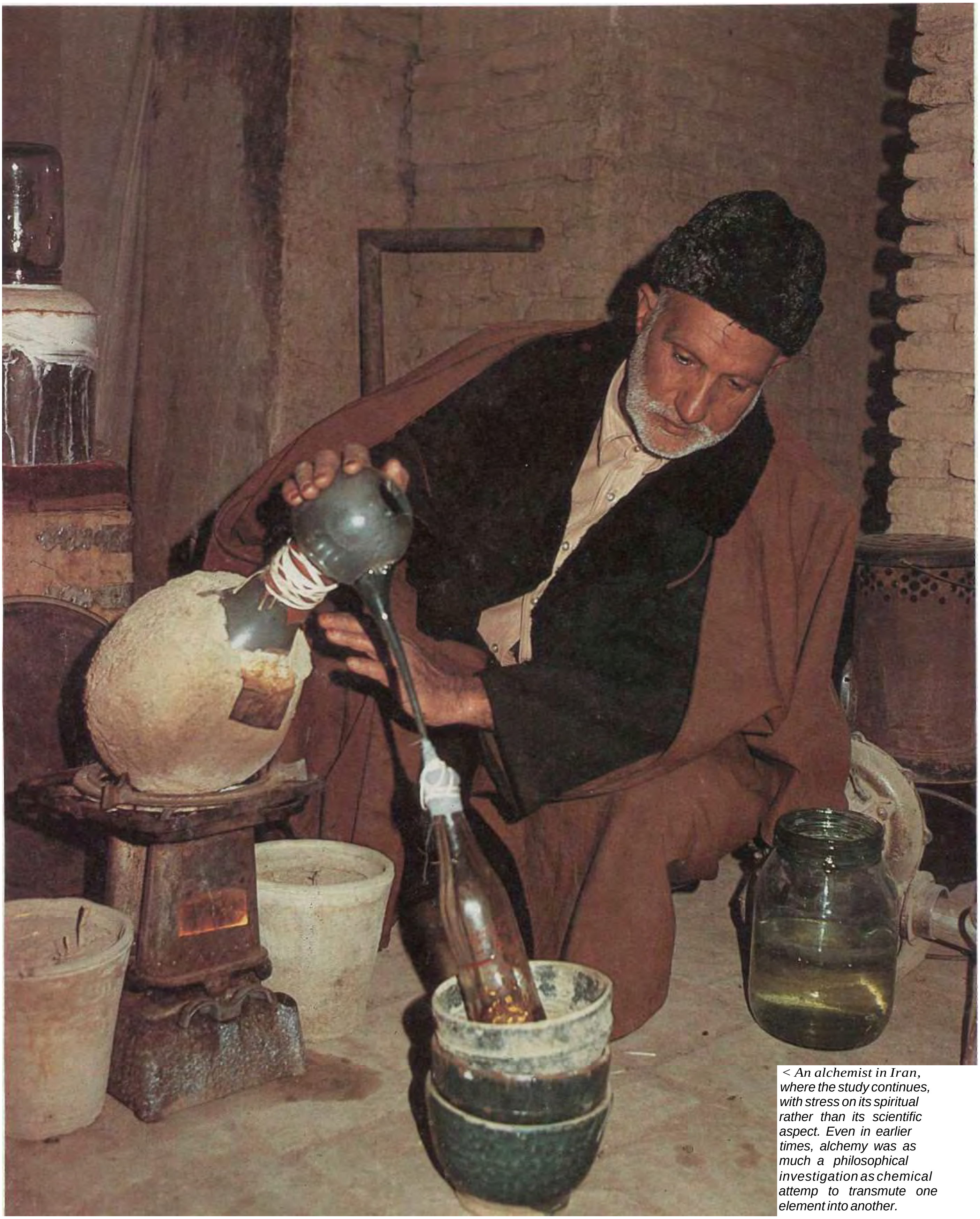
'. .',.,,.;, .1 rfflW i inuttd
IM5Thc¥ia$*n)
AHiwfan
Uxl«dUi:cr (>.\l I 31'*
Ei«bud

ISBNIWIBWMX

Primal in 5pna bj II FIYIIMMT, S A

Contents

:



< An alchemist in Iran, where the study continues, with stress on its spiritual rather than its scientific aspect. Even in earlier times, alchemy was as much a philosophical investigation as chemical attemp to transmute one element into another.

Studying the Material World

1

The ancient view of matter...Greek science...Islamic astronomy, physics and alchemy...Medieval science... Dalton and modern atomism...Physics and chemistry in the 19th century..Modern physics and chemistry... PERSPECTIVE...Greek atomism...Chinese science... What do physicists and chemists do?

The earliest efforts to understand the nature of the physical world around us began several thousand years ago. By the time of the ancient Greeks, over 2,000 years ago, these attempts at explanation had become both complex and sophisticated. They were characterized by the desire to find a single explanation which could be applied to all happenings in the physical world. For example, the description of the world that received most support supposed the existence of four primary chemical elements - earth, water, air and fire. This list may look odd to us but we should see it as something like the modern division of substances into solids, liquids and gases (-> pages 25-34). These four elements were considered to have particular places where they were naturally at rest. The earth, preferentially accumulated at, or below, the Earth's surface; the water came next, lying on top of the Earth's surface; air formed a layer of atmosphere above the surface; and, finally, a layer of fire surrounded the atmosphere. This layering of the elements was invoked to explain how things moved on Earth. A stone thrown into the air fell back to the Earth's surface because that was its natural resting-place; flames leapt upwards in order to reach their natural home at the top of the atmosphere, and so on.

Greek philosophers set the scene for later studies of the material world by distinguishing between different types of theories of matter. The Greeks pointed out that two explanations are feasible. The first supposes that matter is continuous; so that it is always possible to chop up a lump of material into smaller and smaller pieces. The other theory supposes that matter consists of many small indivisible particles clumped together; so that chopping up a lump of matter must stop once it has reached the size of these particles.

The four humors

The chemical elements could combine to create new substances - in particular, they formed the "humors". Each individual human being contained a mixture of four humors, made up from the four elements, and the balance of these humors determined the individual's nature. This theory is still invoked today when we say someone is in a "good humor". Indeed, some of the Greek technical terms are still used: "melancholy" is simply the term for "black bile", one of the four humors. So the chemical elements of the ancient Greeks were involved in determining motion, a fundamental part of physics, and in determining human characteristics, an area now referred to as physiology and biochemistry. The Classical world did not distinguish between physics and chemistry, but saw all of what we would now call "science" as an integrated whole, known as natural philosophy; by the end of the period, however, a distinction between the two areas of study was beginning to emerge as practical studies in alchemy developed that field into a separate area of knowledge.

The Greek view of matter

The debate on whether matter was continuous or made up of discrete elements began with the earliest known Greek thinker, Thales (c.624-c.547 BC), who asserted that all matter was made of water. By "water" he meant some kind of fluid with no distinctive shape or color. Subsequently, Anaximenes (c.570 BC) suggested that this basic substance was actually air. Again, by "air" he meant not just the material making up our atmosphere, but an immaterial substance which breathed life into the universe. These early views led to the popular Greek picture of matter described first by Empedocles (c.500-c.430 BC), where there were four elements - earth, water, air and fire. All these proposals implied that matter is continuous.

The opposing view appeared later, beginning with the little-known Leucippos (c.474 BC) and fully expounded by his pupil Democritos (c.460-c.400 BC). This saw matter as consisting of solid "atoms" (the word means "indivisible") with empty space between them. The idea of empty space was, in its way, as great an innovation as atoms; for continuous matter left no gaps. Both views flourished in ancient Greece, but a belief in continuous matter was much commoner. The debate restarted in 17th-century Europe, still on the basis of the early Greek speculations, but this time it finally led to an acceptance of atomic matter (-> pages 8-9).

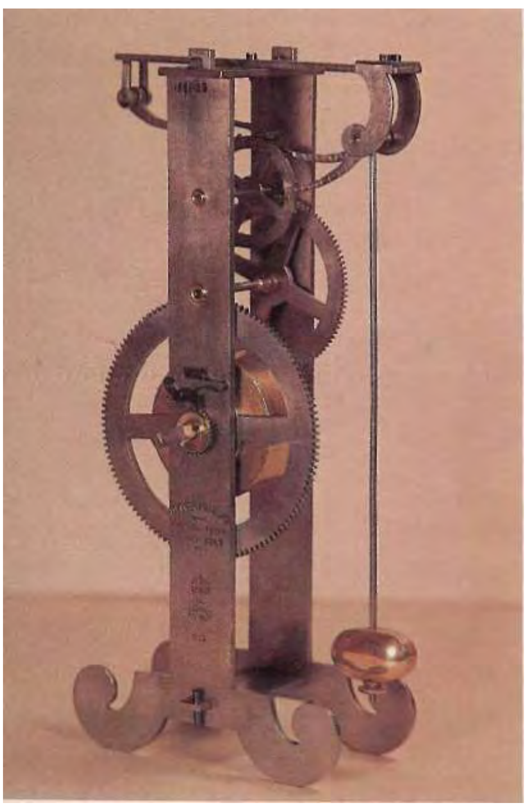
• Much ancient study was devoted to the movements of the Sun, Moon and planets. Monuments such as Stonehenge in southern Britain were used as observatories. Here a partial eclipse of the Moon is seen above Stonehenge.

Early Chinese physics and chemistry

The early Chinese view of the world differed in important respects from the Greek. The Chinese saw the world as a living organism, whereas the Greeks saw it in mechanical terms. In some ways this made little difference. For example, the Greeks concluded that all matter was made of four elements; the Chinese supposed there were five - water, earth, metal, wood and fire. The Chinese, like most Greeks, believed matter to be continuous. Perhaps their picture of the world as an organism prevented them from thinking of the alternative atomic theory, unlike the Greeks.

The Chinese led the world for many centuries in practical physics and chemistry. Their knowledge of magnetism advanced rapidly. They learnt at an early date how to magnetize iron by first heating it, and then letting it cool whilst held in a north-south direction (-> page 47). They realized, 700 years before Western scientists, that magnetic north and south do not coincide with terrestrial north and south. In chemistry, too, practical knowledge was ahead. Thus, experiments seeking for the elixir of life led instead to the discovery that a mixture of saltpetre, charcoal and sulfur formed the potent explosive known as gunpowder.

Why then, with this practical lead, did modern physics and chemistry not originate in China? Factors that have been suggested include the limitations of Chinese mathematics, the nature of the society, and even the structure of the language.



A A reconstruction of Galileo's pendulum clock. The development of accurate clocks enabled scientific measurement, and allowed him to develop the study of forces and motion, initiating modern physics (-> page 11).

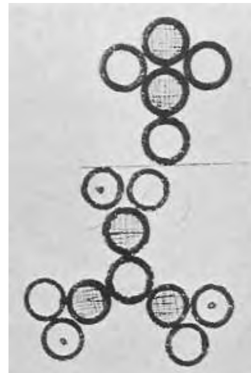
The division between physics and chemistry

One of the great problems in discussions of motion was to try and explain how the Sun, Moon and planets moved across the sky. This question had been enthusiastically attacked by the ancient Greeks, and their work was followed up by the Arabs, but in both cases on the assumption that all these bodies moved round a stationary Earth. The concentration on astronomical motions reduced interest in the link between physics and chemistry. The Greeks and Arabs believed that the heavens were made of a fifth element - labelled the "aether" - which had nothing in common with the terrestrial elements. Consequently, motions in the heavens could not be explained in terms of motions on the Earth; so study of these motions held little of consequence for the relationship between physics and chemistry.

At the same time, a form of chemistry arose which also diverted attention away from the link with physics. Called alchemy, it emphasized practical activity along with a diffuse theory, typically expressed in symbolic terms. Though alchemy first appeared in the late classical world notably in Alexandria, now in Egypt, it flourished particularly amongst the Arabs. A major aim was to transmute one metal into another, especially to turn "baser" metals into gold. Alchemists thought this could be done by finding an appropriate substance - often called the "philosopher's stone" - which would induce the change.

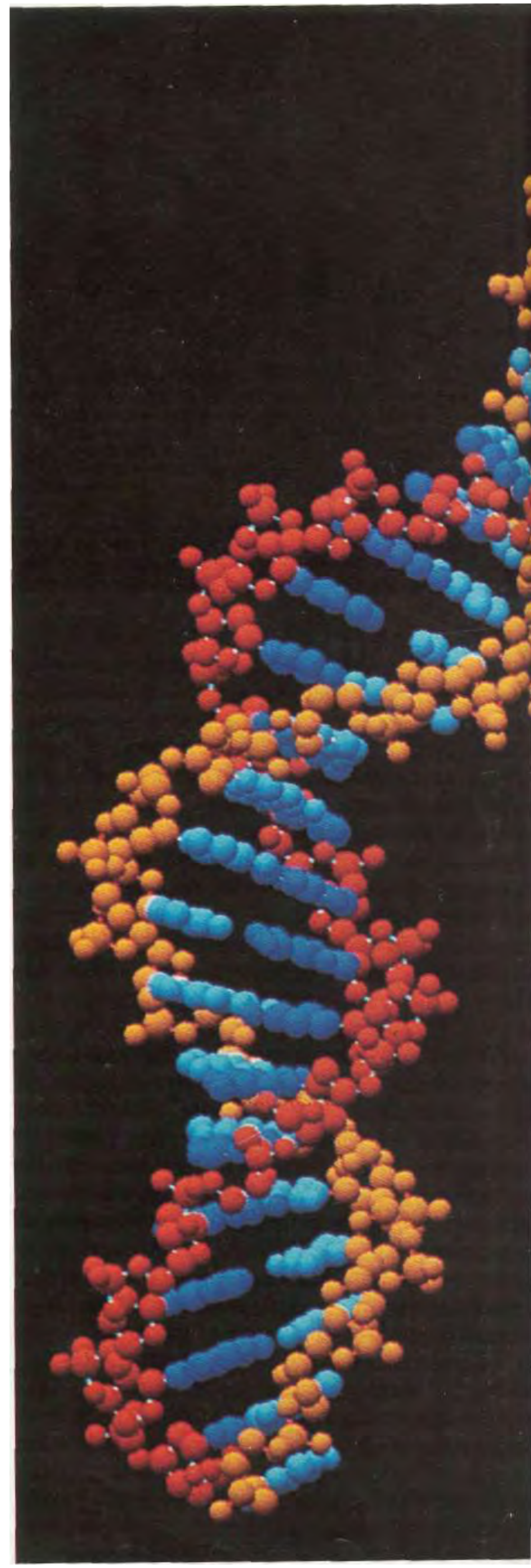
Over the centuries, Arabic studies led to a number of practical developments in physics and chemistry, but retained much the same theoretical framework as the Greeks. From AD 1100 onwards, scholars in western Europe began to translate and study the Greek texts preserved by the Arabs, along with the developments made by the Arabs themselves. As the Arab world became gradually less interested in science, the Western world caught up and, by the 16th century, had reached the point where it could advance beyond either the Greeks or Arabs. The first breakthrough was in astronomy. A Polish cleric, Nicolaus Copernicus (1473-1543), worked out how the motions of the heavens could be explained if the Earth moved round the Sun, rather than *vice versa.* His initiative led over the next 150 years to an explanation of planetary motions that is still basically accepted today. This explanation showed that motions in the heavens and on the Earth were not basically different, as had been previously supposed. It also overthrew the old idea of a connection between the chemical elements and the nature of motion. A division between physics and chemistry therefore remained unbridged, as physics remained linked to astronomy and chemistry to alchemy. The English scientist Isaac Newton (1642-1726), for example, was not only one of the greatest mathematicians and physicists of all time, he was also an enthusiastic alchemist. Yet he seems to have made little connection between these activities.

One step in the 17th century which held some hope for renewing links between physics and chemistry was the fresh interest in an atomic theory. The idea that all matter was made up of tiny, invisible particles called "atoms" originated with the ancient Greeks, but has always been less popular than the belief in four elements. It was now revived, with the suggestion that the various materials in the world might all be formed from atoms grouping together in various ways. This sounds a very modern explanation, but it was not very useful in the 17th century. Atoms could not be studied, or their properties determined, with the equipment then available. So physics and chemistry continued to develop along their own lines.



< ^ John Dalton was the first chemist to show molecules as compounds of elements arranged in a particular manner. His formulae for organic acids (1810-15) are shown here.

• A modern computer graphic illustration of part of the DNA molecule, which contains the genetic code.

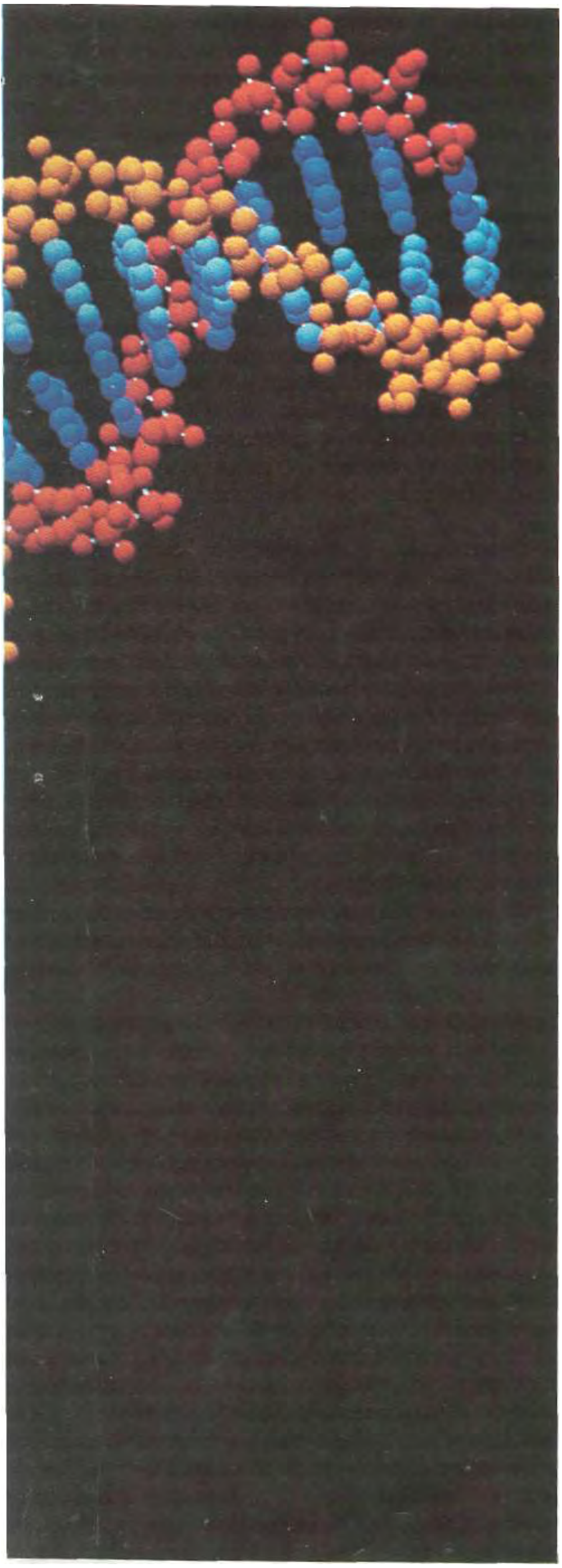


The 19th century

Up to the 18th century, physics had progressed more rapidly than chemistry, but now chemistry moved ahead. The theories of alchemy were rejected, but its concern in practical experiments was pursued vigorously. One area of particular concern was the analysis of gases. It became clear that the old element "air" actually consisted of a mixture of gases; other gases, not present in the atmosphere led to two major developments. In the first place, the Frenchman, Antoine Lavoisier (1743-1794) introduced the modern definition of a chemical element and the modern idea of elements combining to form a variety of chemical compounds. Secondly, John Dalton (1766-1844) in England and Amadeo Avogadro (1776-1856) in Italy showed that elements combined in simple proportions by weight, as would be expected if matter was made up of atoms.

This concept of chemical compounds as a series of atoms linked together led to one of the basic scientific advances of the 19th century. Each atom was assigned a certain number of bonds - now called "valence" bonds - by which it could attach itself to other atoms. The results of chemical analysis could be interpreted in terms of valences, and the theory also formed the basis for the synthesis of new compounds. Knowledge of chemical bonds improved throughout the century. For example, the carbon atom was assigned four valence bonds. From studying the properties of carbon compounds, chemists worked out where in space these bonds pointed relative to each other. The spatial picture they derived was found to explain quite unrelated physical observations. It was also known that some properties of light were changed when it was passed through certain organic compounds. The chemists' explanation of carbon-atom bonding proved capable of explaining why the light was changed. In these instances chemistry provided a better insight into the nature of matter than physics could.

To most 19th-century physicists, atoms were little more than tiny billiard balls. Chemists recognized that atoms must be more complex than that, but could not, themselves, provide a better description. It



was the physicists who made the important breakthrough. Again, it came from the study of gases - in this case, from examining the passage of electricity through rarified gases. Experiments by the British physicist J. J. Thomson (1856-1940) showed that electrical "cathode rays" in gases seemed to consist of sub-atomic particles, which gave some insight into the nature of atoms. Thomson discovered that atoms contained particles - which he labeled "electrons" - with a low mass and a negative electrical charge (-> page 69). Not long afterwards, the New Zealander Ernest Rutherford (1871-1937) deduced that atoms consisted of a cloud of electrons circling round a much more massive positively-charged nucleus (-> page 79).

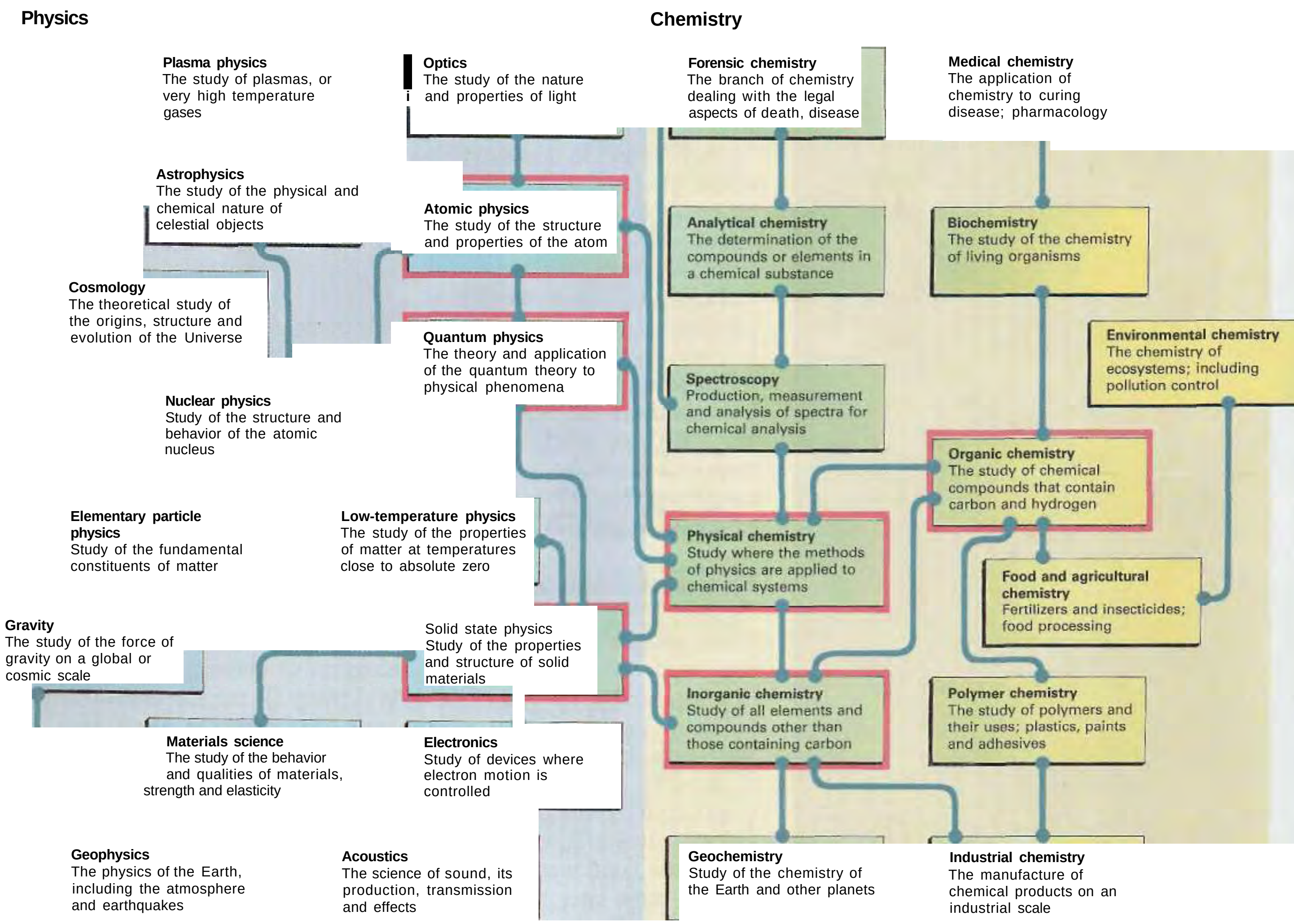
These were startling developments, but it was the next step that had the most impact on chemists - the explanation, "quantum mechanics" began with Niels Bohr (1885-1962) just before World War I, but reached a stage where it was useful in the 1920s. Quantum mechanics showed how electrons in different atoms could interact, so linking the atoms together. Now the valence bonds of the chemists could be explained in terms of the physicists' atom (-> page 87).

Physics, chemistry and industry

By the 1920s the theoretical link between physics and chemistry was firmly established. But the practical applications of the two subjects continued on separate paths. A recognizable chemical industry had first appeared at the end of the 18th century. It remained small-scale for many years, and was mainly concerned with the production of simple chemicals, such as household soda (NaOH). In the latter part of the 19th century, attention turned to the production of organic compounds (containing carbon). The successful synthesis of new artificial dyestuffs led to a rapid growth of the chemical industry, which has continued ever since. An industry based on research in physics came later than in chemistry; but, by the end of the 19th century, earlier studies of electricity and magnetism had led to thriving industries in electrical engineering and communications. These physics-based industries had little in common with the chemical industry, and the gap was not bridged by any major developments in the first half of the 20th century.

The position has changed drastically in recent decades. Science, industry and defense have become intermeshed in a variety of ways, several of which involve joint activity in physics and chemistry. A good example concerns the Earth's upper atmosphere. This is a region of considerable importance, both for space activities and for military purposes. How it can be used depends on the properties of the gases present, and determining these has led to co-operative investigations of the region by physicists and chemists. However, the most revealing example of interdependence is molecular biology. The nature of biological materials has long been studied by applying various physical and chemical techniques, the most important being their interaction with X-rays. Results initially came slowly because of the complexity of biological compounds. But researchers, mainly in Britain and the United States, gradually pieced together information about the nature of biological molecules. The most significant advance was made in 1953, when Francis Crick (b.1916) and James Watson (b.1928) were able to describe the structure of the basic genetic material, DNA. From that work has come the new "biotechnology" industry. Today, the ancient Greeks' belief that these three branches of science are linked has been vindicated, but in a way far beyond their envisaging.

See also
Forces, Energy and Motion 11-20
Atoms and Elements 65-72
Studying the Nucleus 79-86



The range of physics and chemistry

Modern physicists and chemists can apply their skills to almost any area of science or technology. This is not too surprising. Questions involving physics and chemistry are basic to almost any attempt at understanding the world around us. So there are scientists who study the physics and chemistry of stars and planets, while others examine the physics and chemistry of plants and animals. The list is endless.

Physics has traditionally been divided into such categories as sound, heat, light, and so on. These divisions hardly suggest the complexity of modern physics, but do hint at the opportunities for applying physics. For example, the design of musical instruments now requires a detailed knowledge of sound. So does the design of music centers, and these also use the products of the huge new microelectronics industry, which is based on electromagnetism and solid-state physics. Physicists in this industry are concerned with applications varying from computers to biosensors (to detect the physical characteristics of living organisms). Electromagnetism figures in most modern forms of communication, and physicists are concerned with improvements to telephones, radio and television. Lasers have been developed for purposes ranging from communication at one end to medicine at the other (where they are controlled by medical physicists). Lasers also appear in one of the most publicized employment areas of modern physics-the attempts to gain new sources of energy from atoms, as via fusion.

Chemistry, too, has its traditional divisions - into physical, inorganic and organic - but, as in physics, the boundaries are blurred nowadays, just as the boundaries between physics and chemistry themselves are increasingly doubtful. Chemists, like physicists, are often concerned with sources of energy. The oil industry, for example, employs chemists on tasks ranging from the discovery of oil to its use in internal combustion engines.

The pollution caused by such engines is monitored by other chemists, for environmental chemistry has expanded greatly in recent decades. Pollution studies often involve looking for small amounts of chemical, a problem shared by forensic scientists as they try to help the police. Much of this work consists of analysis - finding what substances are there - but many chemists are more concerned with the synthesis of new compounds. Vast amounts of time and money are spent on this in the pharmaceutical industry. Finally, physicists and chemists must think of the future of their subjects: so many are employed in some area of teaching.

▲ Together physics and chemistry provide a framework of interlinked subject areas that are used to explain matter, energy and the Universe. Physics has the wider span, encompassing the smallest subatomic particle at one extreme, and the infinity of the known Universe at the other. Chemistry, however, may limit itself to the level of atoms and molecules but these are the building blocks of all matter. In some areas, in the center of the diagram, physicists and chemists may be studying the same phenomena, but approaching them from different angles or asking different questions. Most of the disciplines in the boxes of this diagram emerged only in the past 50 years.

Forces, Energy and Motion

Why do objects move?...Newton's laws of motion... Friction...Energy at work...Conversion of energy... Oscillating systems...PERSPECTIVE... Vectors, velocity and acceleration...Circular motion...Gravity...Newton and the apple... The tides... The physics of pool... Defining work...Resonance

Imagine a ball being hit by a stick like a golf club. The impact producing the movement is obvious, and the ball eventually stops rolling. Ancient Greek philosophers were puzzled by such situations because they could see no reason for the ball to continue moving after contact with the stick has been broken. Aristotle (384-322 BC) believed the medium through which the ball moves transmits thrust to the ball.

Eventually the Italian scientist Galileo Galilei (1564-1642) concluded that the problem was being considered from the wrong viewpoint. He argued that constant motion in a straight line is as unexceptional a condition as being stationary, but the continual presence of friction (^ page 15) on moving objects conceals this. Without friction the ball would roll in a straight line forever, unless its direction is changed by hitting another object. It is therefore only *changes* in motion that deserve particular consideration.

Velocity and acceleration

Physicists distinguish between the concepts of speed and velocity. Speed indicates the distance covered by a body in a given period of time, irrespective of the direction it is moving. It may be measured in meters per second, for example. Velocity, on the other hand, is a so-called "vector" quantity: that is, a quantity that requires direction as well as magnitude. Two ships that travel equal distances in equal times have the same speed, but they have the same velocity only if they move in the same direction. Because directions are involved, adding velocities and other vectors requires special techniques. These involve drawing parallelograms in which each line represents the distance covered and the direction of each vector.

Acceleration (which is another vector quantity) is defined as the change in velocity per second, measured in meters/second^2 (m/s^2). A satellite in circular orbit will be traveling with constant speed, but its direction is continually changing. As a result, its velocity is similarly changing, and so it must have an acceleration. This acceleration is towards the center of the orbit, and is caused by gravity (tpage 14).

T *Motion is no more unusual than standing still; it is changes in motion that involve an external influence. When a horse slows down abruptly, the rider tends to continue in the same state of motion, and tumbles over the top.*

Galileo also considered the motion of falling bodies, and showed that any two objects in free fall at the same place above the Earth's surface have the same acceleration. He deduced the basic relationships of dynamics, showing that the velocity of a uniformly accelerating body increases in proportion to the time, while the distance traveled is proportional to the time squared. Why all falling bodies should have the same acceleration was an unanswered question.

When the English scientist Isaac Newton (1642-1727) came to consider this problem, he set down three "laws of motion" as a foundation upon which to build his revolutionary theory of gravitation.

Law 1 stated that "a body will continue at rest, or in uniform motion in a straight line unless acted upon by a resultant force". Newton introduced the idea of "mass", or inertia, as a measure of a body's reluctance to start or stop moving.

In his second law ("the rale of change of momentum of a body is proportional to the resultant force on the body, and takes place in the direction of that force"), Newton attempted to describe the change in motion that a body would experience under the action of a resultant force. He introduced the quantity "momentum", the product of mass times velocity. In cases where the mass of the body is constant, this second law is stated simply as "force equals mass times acceleration".

Law 3 states that "if a body A experiences a force due to the action of a body B, then body B will experience an equal force due to body A, but in the opposite direction." Newton illustrated his third law through the example of a horse pulling a stone tied by a rope. While the stone experiences a force forwards, the horse experiences a force backwards. The tension in the rope acts equally to move the stone and to impede the movement of the horse.

A consequence of Newton's second and third laws is that when two objects collide with no external forces acting upon them, the total momentum before the collision is equal to the total momentum after the collision. This is the "conservation of linear momentum", and is of great value in analyzing collisions or interactions on any scale. For example, when a gun fires, the momentum of its recoil is equal and opposite to the momentum of the bullet, adding to a total momentum of zero - the same as before firing.

A *Once hit, an ice hockey puck shoots in a straight line, demonstrating Newton's first law of motion. According to his second law, the heavier an object, the greater the force needed to set it moving, as anyone knows who has tried to push or pull (right) a truck. Newton's third law equates action (here the upward pull of the athlete's muscles) with reaction (the downward force of the car's weight).*

• These people flying rounds roundabout do not travel in a straight line because they feel a centripetal force, acting toward the center of their circular path. This force is the net result of the weight of the chair and body, acting downward, and the tension in the wires.

Circular motion

An object such as a seat on a fairground roundabout, traveling in a circle, can appear to be moving uniformly. However, its velocity is continually changing. To understand why, recall that velocity is a vector quantity, with a direction as well as a magnitude. At any point in time the velocity of the seat is in fact in the direction of the tangent to the circle at the roundabout's position. As the seat moves, this direction, and hence the velocity, changes. According to Newton's first law the seat must therefore be subject to a force and, indeed, this force is applied continually to the seat via the chain that holds it to the roundabout. If the chain were to break and the force it provides were thus suddenly interrupted, the seat would fly away in a straight line, as Newton's first law dictates.

Any force that produces circular motion of this kind is called a "centripetal force". It acts towards the center of the circle, and therefore at right angles to the motion round the circle. The size of the force is equal to the mass of the object multiplied by the square of the speed and divided by the radius of the circle. Here, the speed is the magnitude of the velocity.

Any object moving on a curved path or rotating on its own axis has an "angular speed". This is the angle the object travels through, with respect to the center of its motion, during a unit of time. An object traveling uniformly in a circle, like the roundabout seat, has a constant angular speed, although its velocity is changing all the time.

Objects with angular speed have "angular momentum", directly analogous to the "linear momentum" of objects moving in straight lines. Angular momentum is equal to mass multiplied by linear speed multiplied by the radius of the motion. In any system, the total angular momentum must be conserved if the system does not experience a turning force, or torque. So if, for instance, the radius decreases, the velocity increases provided the mass remains the same. This is why, for example, a figure skater spins slower when she stretches out her arms horizontally and faster when she pulls them in.

The concept of gravity enabled scientists to describe the orbits of the planets, the rhythms of the tides, falling objects and many other phenomena

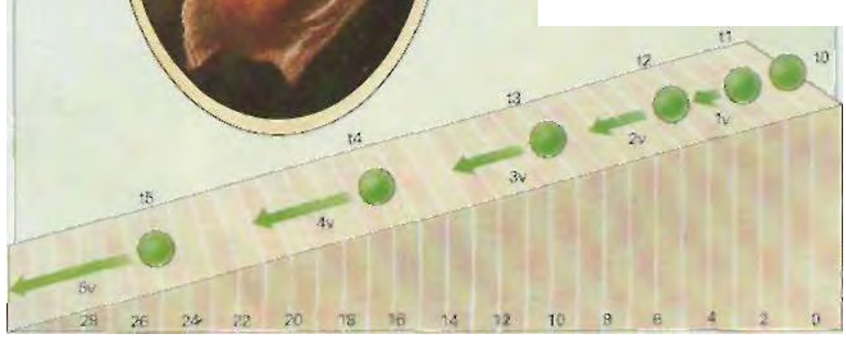
Gravity

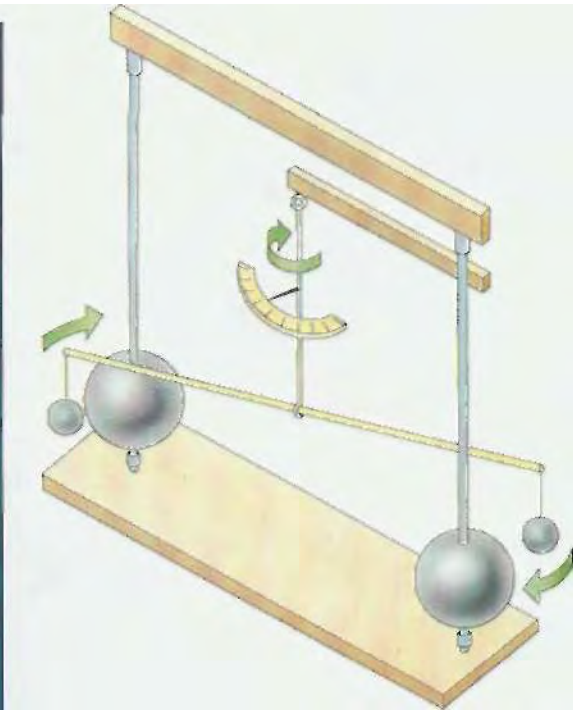
Gravity is the most obvious of nature's forces (p page 105). It keeps us on the ground, and it controls the behavior of the Universe. The structure and motion of the planets, stars and galaxies are all determined by gravity.

Newton was the first to realize that all bodies with mass attract each other. He showed that the force of attraction between two bodies is proportional to the product of their masses times a constant, and inversely proportional to the square of their distance apart.

The constant here is called the universal gravitational constant. It is usually denoted by G and is equal to 6-673 x 7 0 " newton meters ^{2}per kilogram ^{2}. In proclaiming this a universal constant, Newton was assuming that the heavenly bodies - the Moon and the stars - obey the same rules as objects here on Earth. This was a revolutionary advance. From the time of the Greek philosopher Aristotle (384-322 BCj, people had believed that earthly and heavenly objects obeyed different laws (4 page 7). After Newton, however, physics could take the Universe as its laboratory; and his point of view remained unchallenged until the final years of the 19th century ($ page 42).

-« • Galileo is well known for reputedly dropping objects of different masses from the tower of Pisa. An experiment he did perform involved rolling steel balls down a gently sloping plank and measuring the distances moved in equal intervals of time, marked by a water clock. This showed that the velocity increased uniformly with time as the ball moved down the slope under the force of gravity.





< Free-fall parachutists experience a force due to air resistance that is equal and opposite to the force due to gravity. Thus, in accordance with Newton's first law of motion, they fall at a constant velocity.

V Fishing boats lie stranded on the sands around a harbor at low tide, as the seas respond to the changing gravitational pull of the Moon across the Earth's diameter.

A The English physicist Henry Cavendish (1731-1810) made the first measurements of the gravitational constant, using a "torsion balance". Two small balls were attached to the ends of a bar suspended at its center by a wire. Large balls held at either end, but on opposite sides of the bar, attracted the small balls through the gravitational force between them, and made the bar twist.

"God said let Newton be, and all was light"

Isaac Newton was born in January 1643 in Woolsthorpe, Lincolnshire. As a schoolboy he was fascinated by mechanical devices and he went up to Cambridge University in 1660, graduating in 1665. When bubonic plague reached Cambridge in 1665 he returned to his mother's farm. The enforced rest left him free to develop his ideas on the law of gravitation which he published 20 years later, in his book "Principia Mathematical At the same time he started a series of optical experiments and discovered, among other things, that white light is a mixture of colors ($ page 38).

Newton was absent-minded and sensitive to criticism. He conducted an international dispute with the German mathematician Gottfried Wilhelm Leibniz (1646-1716) as to who had first discovered calculus. Nearer to home, he quarreled for years with the British physicist Robert Hooke (1635-1703). Hooke claimed that Newton had stolen some of his ideas and put them in the "Principia ". Newton was finally forced to include a short passage acknowledging that Hooke and others had reached certain conclusions which he was now explaining in greater detail. These quarrels infuriated Newton, and contributed to his nervous breakdown in 1692.

Much of Newton's life was spent in trying to manufacture gold and in speculating on theology, yet he was honored and respected as few scientists have been before or since.

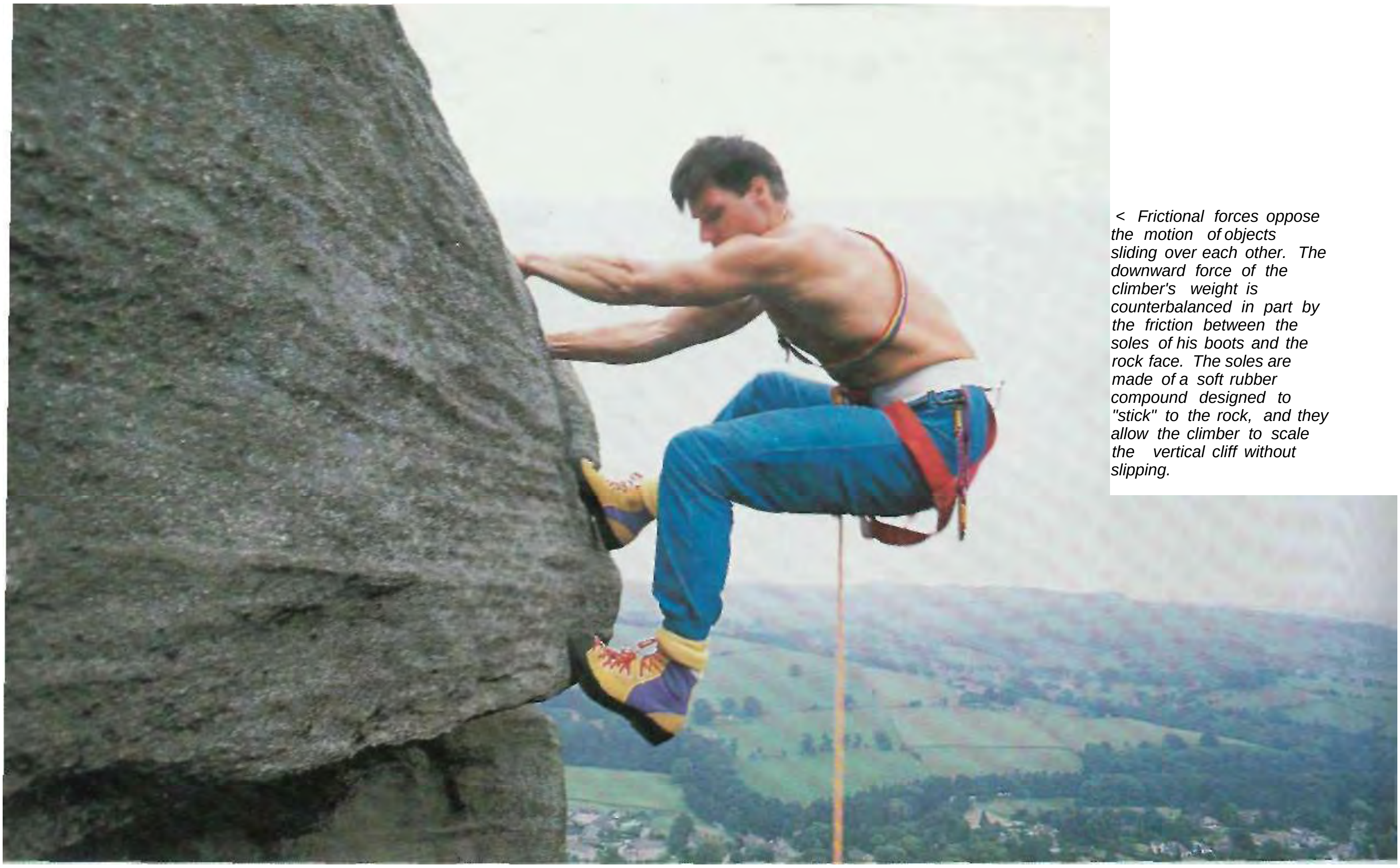
Gravity and the tides

The Earth and the Moon rotate about their common center of mass (the point where an outsider would consider all the mass of the system to be concentrated). Because the mass of the Earth is so much greater than that of the Moon, the center of mass is much closer to the Earth than to the Moon.

Newton showed that bodies move in straight lines at constant speed unless a force acts upon them. Thus there must be a force that keeps the Earth orbiting around the center of mass of the Earth-Moon system. This force, which is centripetal, is provided by the gravitational attraction of the Moon, and it is just the right size to keep the center of the Earth orbiting about the center of mass.

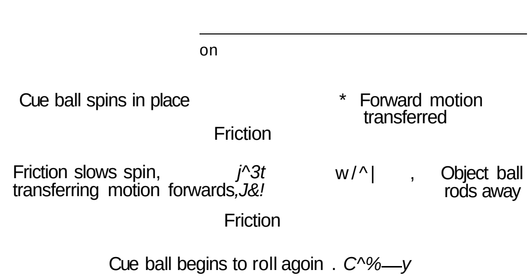
The Moon's gravitational force decreases as the distance from the Moon increases. For points on the Earth closer to the Moon than the Earth's center, the gravitational force is larger than required for the orbital motion. Here the Earth is stretched towards the Moon. The seas, being free to move, bulge towards the Moon. For points farther from the Moon than the Earth's center, the gravitational force is weaker than required and the seas bulge out away from the Moon. The Earth spins on its axis, rotating under these bulges which sweep over the surface of the Earth, causing two high tides each day.

The gravitational pull of the Sun also causes tides, but the Sun is so much farther from the Earth than the Moon that its gravitational pull changes less across the Earth's diameter. The tides are largest (spring tides) when the Sun, Moon and Earth reinforce each other, and weakest (neap tides) when the three bodies are 90" out of line and the tidal effects of the Sun and Moon tend to cancel.



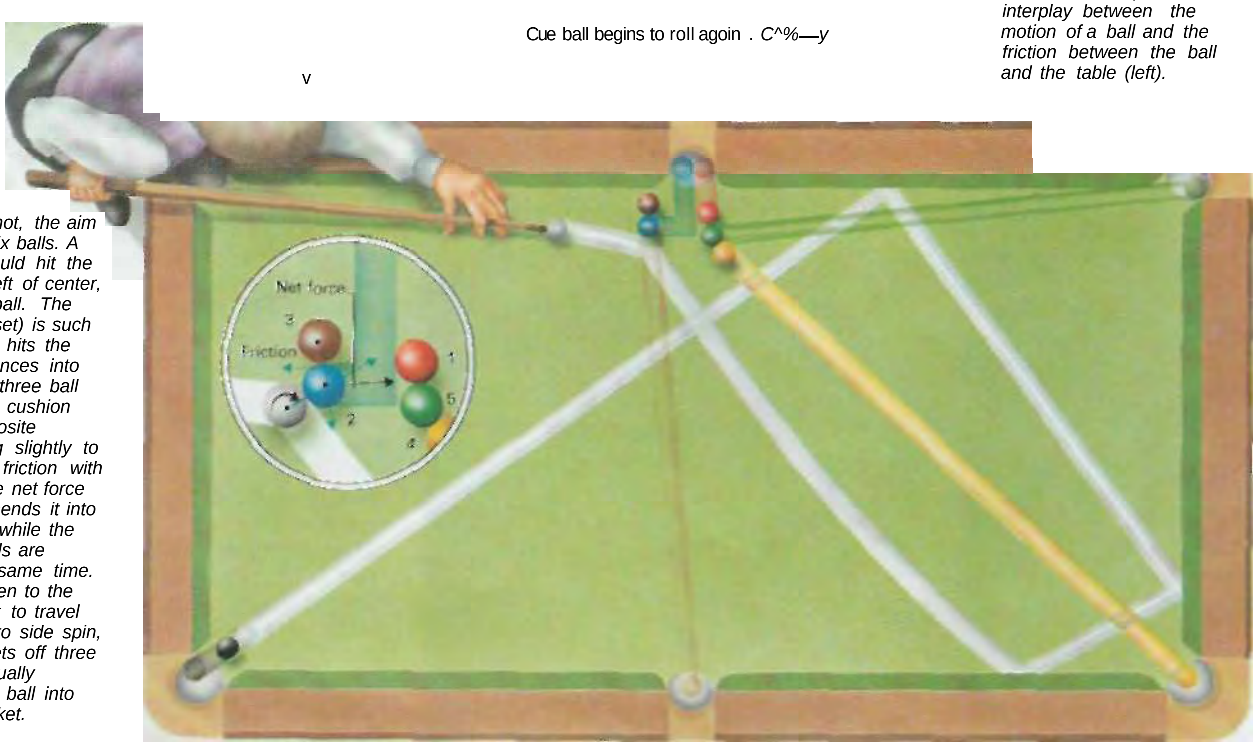
< Frictional forces oppose the motion of objects sliding over each other. The downward force of the climber's weight is counterbalanced in part by the friction between the soles of his boots and the rock face. The soles are made of a soft rubber compound designed to "stick" to the rock, and they allow the climber to scale the vertical cliff without slipping.

The follow shot



< In a game of pool a cue ball hit slightly above center (for left) is given "top spin", rotating in the direction of its motion; cueing below the center results in "backspin". Positioning the cue to left or right imparts "side spin ", which allows the cue ball to swerve in the correct conditions. In detail, shots depend on the interplay between the motion of a ball and the friction between the ball and the table (left).

A trick shot



> In this pool shot, the aim is to pocket all six balls. A skilled player would hit the cue ball above left of center, toward the two ball. The net force (see inset) is such that the two ball hits the five ball and bounces into the pocket. The three ball ricochets off the cushion toward the opposite pocket, swerving slightly to the right due to friction with the two ball. The net force on the five ball sends it into the top pocket, while the one and four balls are pocketed at the same time. The top spin given to the cue ball allows it to travel on, curving due to side spin, so that it ricochets off three cushions, eventually knocking the six ball into the bottom pocket.

The physics of pool

The laws of motion are often described in terms of the interactions of "billiard balls ", on the assumption that in a two-dimensional plane the momentum and angular displacement of bodies after collision can be calculated simply from their previous velocity and the angle of impact. It is convenient to think of billiard balls as behaving in this manner but in practice their behavior is more complex, being affected by friction.

When a ball moves across a snooker or pool table it has two types of motion. The first is a forward "translational" motion, the second is a rotation about the ball's center. For pure rolling there is a relationship between these two. In other cases skidding occurs at the table surface. This happens, for example, when a ball is hit cen traliy by a cue. Initially the ball moves off without rotating and slides across the table. However, friction between the ball and the table causes the ball both to slow down and to start rotating. When the rotational motion matches the translational motion pure rolling takes over, and the friction decreases correspondingly.

To eliminate this initial skidding the ball must be set moving with the correct amount of initial rotation. This is achieved by striking it slightly above the center. The cushions on the table are set rather higher than the center of the balls for similar reasons. When a rolling ball hits a stationary one, forward movement of the cue ball is transmitted to the object ball. The object ball moves off skidding, because it has been hit centrally. If the balls are smooth there is no significant friction between them and no rotation is transmitted in the impact The cue ball is left instantaneously stationary, but still rotating. The frictional force which slows this rotation also gives the cue ball forward motion (and if strong enough, it may cause the cue ball to follow the object ball into the pocket!).

If the cue ball is still skidding as it makes the collision, the player has some control over the outcome. For example, if the cue ball is not rotating at all and is simply sliding across the table, it will stop dead after collision with the object ball. If, however, it is hit below its center its rotation will be in opposition to its forward motion, and friction will cause it to move backwards after the collision.

If the collision with the object ball is oblique rather than head on, the cue ball does not lose all its translational motion, but moves off in a different direction at reduced speed. The frictional force resisting skidding is now no longer aligned with the direction of movement. As a result, the ball swerves while skidding continues, before eventually moving in a straight line once pure rolling starts. This gives the player some control over the final direction of the cue ball, in anticipation of the next shot.

Similarly the player may swerve the cue ball around an obstacle. By cueing to the right or left of center, the spin produced is across the direction of forward motion. This resulting sideways frictional force at the surface allows the ball to swerve as long as skidding is taking place. These techniques all require that the cue ball has not started to roll; for a typical, firmly struck shot the ball must not have traveled more than about one meter.

Newton was conscious of two types of force. First there are those that involve contact of some kind including friction, tension and compression. Second, there are forces that are able to act across a distance, such as magnetic () page 45) or electrostatic forces (» page 49) and the force that concerned Newton, gravity. Subsequently, scientists began to interpret forces in terms of the interaction between particles, such as the collisions of air molecules at a surface causing air pressure (^ page 25), or the interatomic forces allowing a wire to withstand tension (*• page 27). The concept of a "field" was introduced to explain forces acting at a distance. Today all the apparently different types of force may be accounted for by four fundamental forces (f page 105).

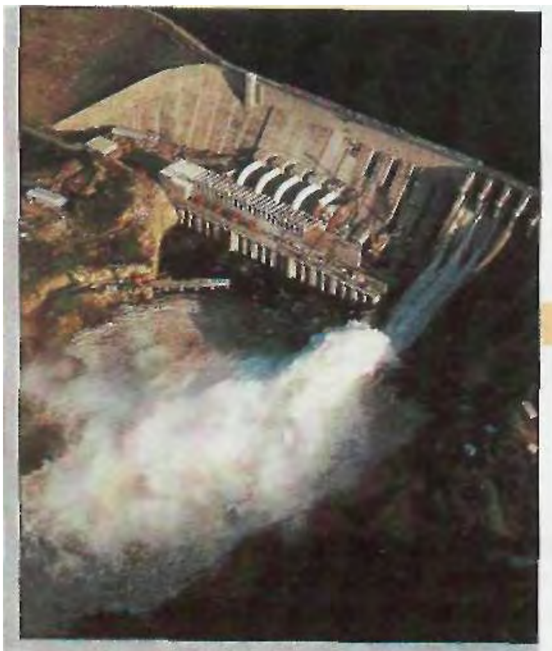
The interplay of forces underlies many physical features of the everyday world. Whenever two surfaces slide over each other, for example, friction has to be considered, even if its effects may be dismissed as negligible. In many circumstances it may be desirable to reduce it as much as possible (by lubrication in engines for example), yet without friction we would not be able to walk, or even stand.

The laws of friction may be demonstrated simply by investigating the force required to pull a block of metal across a horizontal metal surface. The frictional force always acts in the direction that opposes the motion of the block, and can have whatever value is necessary to prevent motion, from zero up to a maximum when sliding occurs. This limiting maximum value depends on the perpendicular force between the block and the surface, but not on the area of contact between the two. It also depends on the nature of the two sliding surfaces. Once the block starts to slip the frictional force usually decreases slightly.

Looking in detail at the surfaces in contact shows that no metal is perfectly smooth. There are only a few points of contact between the block and the surface. Here the local pressure is very high, and interatomic forces (f page 25) tend to bond the two together. For sliding to take place these local joints have to be broken, and this gives rise to the frictional force. As one set of joints is broken others form, in a continuous process. The number of local points of contact does not noticeably rise when the apparent area of contact increases, but does so when there is a larger normal force.

A *Even the highly polished surface of aluminum alloy appears rough through a microscope.*

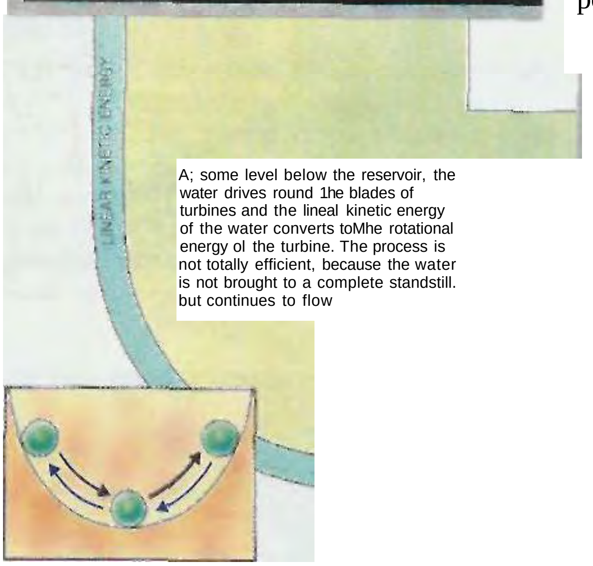
The conservation of energy
A hydroelectric power station taps the store of potential energy that is held in a water reservoir As the water is released, the potential energy is converted to kinetic energy when the water runs downhill



There is a continual interplay between different types of energy. One of the simplest examples is provided by a ball confined to a hollow. If the ball is released at the top of one side of the hollow, it rushes down to the bottom and up the other side, slowly coming to a halt before rushing back down into the hollow and up the first side again. If there were no friction between the ball and the surface, this oscillating movement could continue for ever, but in practice the ball rises up the sides less and less each time until it eventually comes to rest in the base of the hollow.

What exactly is happening to the ball? It gains *kinetic energy* - energy of motion - as it falls into the hollow. The kinetic energy is gained as the ball falls downwards through the Earth's gravitational field. It is lost again as the ball moves upwards, against the gravitational field. The work done by the ball against gravity is defined as the force on the ball (due to gravity) multiplied by the vertical distance moved (that is, the difference between the heights of the top of the slope and the base of the hollow).

The change in energy of the ball is related to the work done - in one sense, an object's energy is its capacity to do work. But this is not the end of the story because once the ball comes to a stop - its kinetic energy is zero - it immediately falls back down the slope. In going up the slope it has gained another kind of energy, known as *gravitational potential energy.* It is a simple matter to show that the potential energy gained equals the kinetic energy lost, while when the ball is at the bottom of the hollow once again, the kinetic energy gained equals the potential energy lost. The total amount of energy remains the same;



A; some level below the reservoir, the water drives round 1he blades of turbines and the lineal kinetic energy of the water converts toMhe rotational energy ol the turbine. The process is not totally efficient, because the water is not brought to a complete standstill. but continues to flow

J
n i

^&W^\iM

A If a ball is released at the top of a hollow, it will roll back and forth, climbing the slope on the opposite side each time, gradually losing height and finally coming to rest at the lowest point. It is continually exchanging potential energy (due to height) for kinetic (due to motion) and vice versa. Gradually the ball loses its energy and comes to rest. Its energy is not destroyed, but rather lost to the system, turned into heat and noise by the action of friction with the surface.

> To a physicist, work takes place whenever a force moves something, or, in other words, when energy is changed to a different form. The greater the distance moved, the more the work done. James Joule was one of the first to appreciate the relationship between heat and mechanical work. The unit of one joule is equivalent to lifting a bag of sugar from one shelf to another in a cupboard; the act of shutting a door might use another five joules.

Once the electricity supply reaches the consumer, the electrical energy is converted to other forms, m particular heat, light and sound — all pervasive at a pop concert. In the home, conversion to mechanical energy occurs in devices from washing machines to lawn-mowers. In cooking, the energy from electricity can fuel chemical changes, as when cakes nse.

one form of energy simply converts into the other, a change that occurs whenever work is done.

The transformation of energy from one kind to another is basic to the machines used in daily life, from simple devices like a can opener to the complex workings of a hydroelectric power station. Even the human body is a machine, continuously converting energy from one kind to another. The body transforms the energy contained in food, for example, to be stored as chemical energy in muscles, before being released as kinetic energy, in a runner, or converting to potential energy in the case of a high-jumper. None of these machines, from the body to a power station, is 100 percent efficient at converting one type of energy to another. In all cases, there are losses.

The principle of the conservation of energy is a fundamental physical law that applies to all kinds of energy: energy cannot be created or destroyed. There are many kinds of energy, but in any process, the total amount of energy always remains the same. As Einstein showed in his theory of relativity (^ page 42), even mass is a form of "frozen" energy, which can be released in nuclear reactions. Electrical, chemical, and nuclear energy are all familiar in our daily lives, as are the forms of energy known better as heat, light and sound. Nuclear energy is used to heat water to drive turbines to produce electricity to heat and light homes; chemical energy released when petrol burns propels many kinds of vehicle. Ultimately most of the energy that is used on Earth derives from the Sun - from the heat that drives the climatic systems, and the light that makes plants grow through photosynthesis.

Defining work

The British scientist James Prescott Joule (1816-1889) was one of the first to appreciate that mechanical work can produce heat. He performed a series of experiments to show the heating effect of work done against friction, including his famous paddle-wheel experiment. For this. Joule used an arrangement of paddles on a central axle, which passed between fixed vanes attached to the walls of a vessel filled with water. As the paddles rotated on the axle, the water became warmed through frictional effects, thus converting the mechanical work done in rotating the paddles into heat, which could be measured through the temperature rise. A system of weights and pulleys allowed Joule to calculate the work done, and so equate work and heat quantitatively.

The modern unit of work done, and therefore of energy, is named in Joule's honor. One joule is the work done in applying a force of one newton through a distance of one meter. On Earth, the gravitational force on a mass of 1kg is 9.8 newtons, so a joule is roughly the energy used (or work done! in lifting 1kg through 0.1m. In terms of heat, the energy required to raise the temperature of 1gm of water through 1 "C is equal to 4.18 joules. Electrical energy, on the other hand, is usually measured in terms of power, or the rate at which energy is flowing. In this respect the unit of power, the watt, is defined as the energy flow of one joule per second.

In the tcbin© house some iy > lost by the turbines h do nrj work aga«isl (fiction as tho.shafts rotate. This . "lost" energy is converted to heat, other losses include the energy •>(the sounds produced »,,; turbines drive generators v.' I convert the -i".etc enerc[1], of the rotating shafts into electrical energy.

The rotation of a turbine shaft in a power station causes a large electromagnet lite rotor — to rotate within a lixed coil, the stator The movements of the electromagnet Induce electric currents' to flow in the stator. thereby converting kinetic energy lo electrical energy. The electromagnet is moved rather than the pickup co'i because it requires relatively low* electric currents to create the magnetic held. The. currents induced in the outer coil are much greater. At this stage losses are about 2 percent.

The electrical energy created by the generator is in the form of alternating current. Large currents at relatively low voltages from the generator are converted to lower currents at higher voltages for transmission. This conversion takes place in transformers, which are very efficient.

WATIQNAI. KINETIC ENERGY

ELECTRICAL ENERGY

Electricity is transmitted by a grid system which links the power stations lo tho industrial and domestic consumers. Overhead transmission lines carry the electricity supply across long distances at high voltages so as to reduce losses that might be caused by electrical resistance in the wires, which dissipate energy as heat

See also
Studying the Material World 5-10
Sound 21-4
Molecules and Matter 25-34
Electricity 49-58
Fundamental Forces 105-10

Resonance

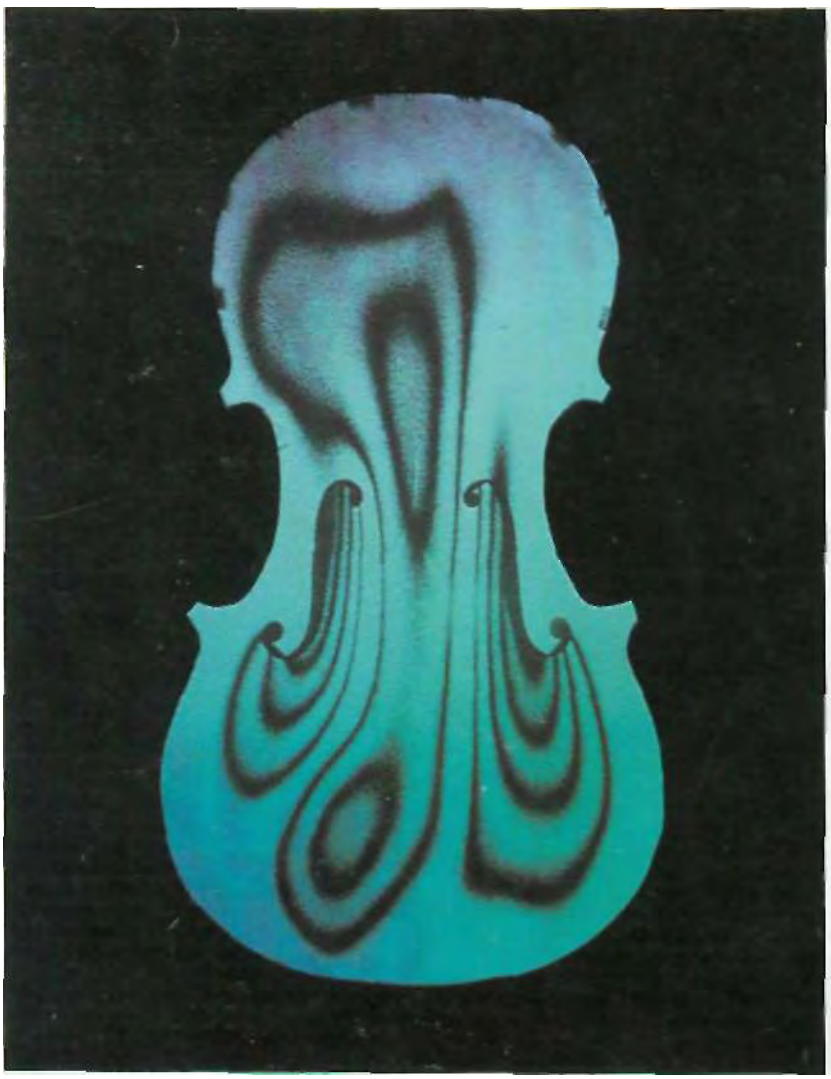
All objects have their own natural frequency of vibration, and when an object is vibrated at this frequency it readily absorbs energy and vibrates through large amplitudes. This condition is known as resonance. It is made use of in musical instruments, in which vibrations are set up deliberately to produce pleasing sounds ($ page 23). But resonance can also be a hazard, as unwanted vibrations can destroy an object. Thus soldiers may be required to break march across certain types of bridge, and it is said that some opera singers can shatter glasses by setting them in resonance with a particular note.

*Resonance is not restricted to mechanical systems. In electronics, a resonant circuit is one in which the frequency response of a capacitor and inductor (} page 64) are matched in such a way that the circuit can pass large alternating currents. Such circuits are used in the transmission of radio waves. In atomic and nuclear physics, resonance occurs when electrons or the nuclei of atoms absorb radiation with a frequency corresponding to a particular transition, as for example in nuclear magnetic resonance f * page 93).*

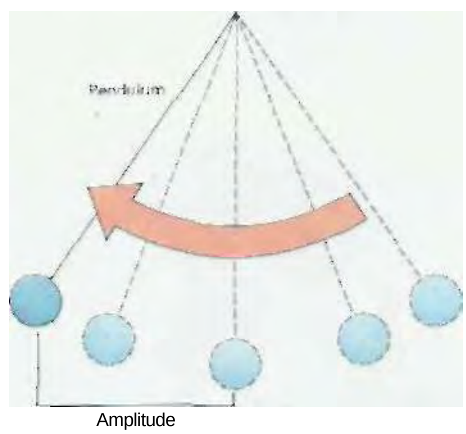
Oscillating systems

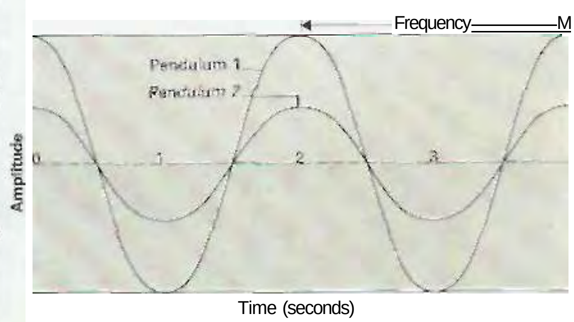
From the motion of the atoms within a molecule to the vibrations of a large engineering structure such as a bridge, oscillations are of great importance. Examples of oscillations such as a mass on a spring, or a pendulum swinging, approximate to "simple harmonic motion". This is an important class of oscillations where the resultant force acting on the moving mass or bob is always proportional to the displacement from the rest position, and directed towards it. Simple harmonic motion (SHM) is important not only because it is common, but because more complex oscillations can be broken down and analyzed in terms of it.

In an oscillating system such as a mass on a spring, there is a continual interchange between the elastic energy stored in the spring and the kinetic energy associated with the movement of the mass. In ideal SHM the period of oscillation is constant regardless of the amplitude of vibration, but il is affected by the elasticity of the spring and the size of the mass. In practical situations energy is lost and so the amplitude decreases. In many cases the motion is deliberately "damped" so that the vibrations die away rapidly. For example, the wheel of a car could oscillate dangerously on the end of the coil spring unless damped by the action of the shock absorber.



Oscillating motion





*•**tin** a violin, the vibrations of the strings pass via the bridge to the body of the instrument. The body has its own modes of vibration - made visible here by interference effects - which resonate with vibrations of the strings. The frequency of these modes is usually constrained to match the frequencies of the strings and gives the violin's tone.*

A The swing of a pendulum bob typifies simple harmonic motion—a regular oscillatory motion that occurs in many physical systems. The angle to the vertical varies between a maximum value (the amplitude) on either side over a definite time period. The time period (frequency) varies only with the length of the string.

Sound

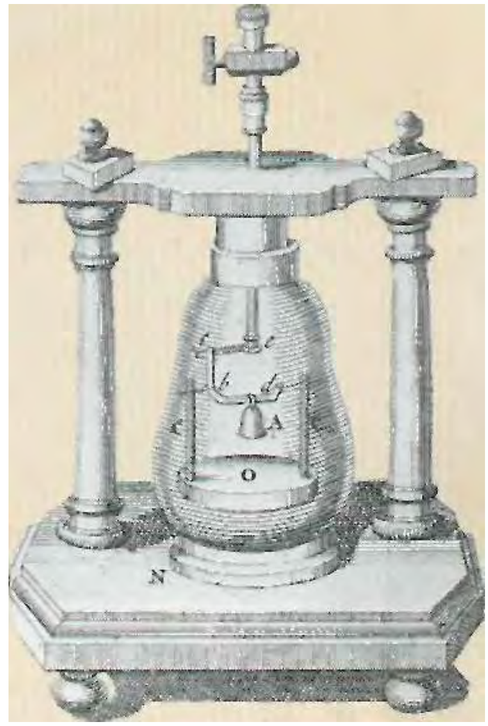
Sound waves...Frt i nicy and wavelength...Diffraction and reflection...PEhoHECTivE„.Loudness and intensity ...Pipes and strings...Sonic booms and the Doppler effect

Some 2,000 years ago the Roman architect Vitruvius (active in the 1st century BC) described the propagation of sounds through the air as like the motion of ripples across the surface of a pond. Vitruvius was largely ignored and it was not until 1,700 years later that the Italian scientist Galileo Galilei (1564-1642) decided for himself that sound is a wave motion, "produced by the vibration of a sonorous body".

A sound wave is a pressure wave and consists of alternating regions of compression and rarefaction. Therefore, unlike a light wave (|-page 61), a sound wave needs a material to travel through.

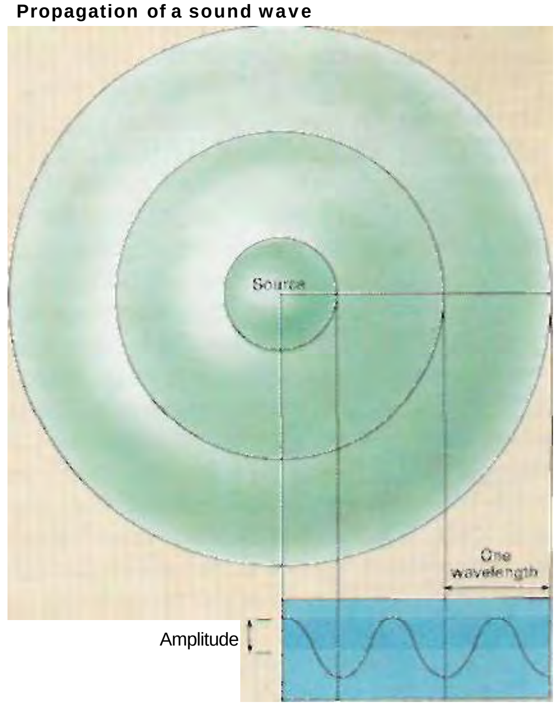
Sound waves are the most familiar example of "longitudinal" waves: waves that vibrate and travel in the same direction. Light, on the other hand, is a "transyerse" wave motion, vibrating at right angles to the direction of travel. The basic characteristics of a sound wave are its "amplitude", its "frequency" and its velocity. The amplitude refers to the size of the pressure variations; the frequency to the number of variations - waves - per second.

The velocity of sound depends on the substance through which it is traveling. Sound moves faster through liquids than gases. In sea water, for instance, the speed of sound is nearly 1,500 meters per second, four times the speed in air, which is a little less than 350 meters per second. In steel, sound travels at 5,000 meters per second. The speed also depends on temperature: the higher the temperature, the greater the velocity. The frequency of a sound wave is related to the "pitch" of the sound: higher notes correspond to higher frequencies, that is more waves per second, or hertz (Hz). Audible frequencies lie in the range 20-20,000 Hz. The inaudible sounds over this higher frequency are referred to as "ultrasonic".



A *Experiments to show that sound waves need a medium such as air to travel through were carried out in the 18th century. Air was pumped from a chamber containing a bell. Without air, the bell no longer made a sound.*

Propagation of a sound wave



A *Sound waves spread out like ripples on a pond, but the ripples are variations in pressure that spread in three dimensions. "Crests" correspond to regions of increased pressure; "troughs" occur where the pressure is lower. Wavelength is the distance between crests; frequency the number of crests that pass a point each second.*

+ *Special photography shows a sound wave from a spark.*

The "intensity" of a sound wave is technically given by ihe square of its amplitude, and it is related to the preccived loudness, albeit in a complicated way. The amplitude of a sound wave represents the pressure change involved, and the smallest pressure variations that can be heard are in the region of 0-00002 pascals (Pa). Human ears are sensitive to a variation in intensity of a factor of a million million.

Echoes and diffraction of sound

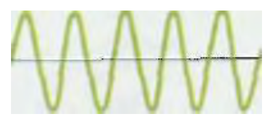
Sound waves demonstrate all the characteristic properties of waves. For example they reflect, refract and diffract just as light waves do. The reflection of sound is a common phenomenon, best known as the familiar echo. In a concert hall echoes can be a nuisance if the hall and its wall coverings are not properly designed, but in other circumstances echoes are vitally important. By timing the reflections of transmitted high-frequency sound waves given off by a sonar device, members of a ship's crew can tell how close their vessel is to the sea bed. And the fact that sound waves are reflected at the boundary between different substances has made ultrasonic sound useful in medical imaging, particularly for an object such as the fetus in a watery environment such as the womb. The refraction, or bending, of sound waves is most apparent at night when sounds often seem louder than during the day. This is because sound can travel further at night, being bent (refracted) back towards the ground by the atmosphere. Refraction occurs when a wave moves into a medium in which its velocity changes. Sound moves faster through warm air, and at night the air near the ground is cooler than the air above it. Sound waves traveling upwards into the warmer air are bent back towards the ground, carrying the sound far along the surface.

Although sound waves propagate basically in stiaight lines, sound can travel round corners - a wave phenomenon known as diffraction. The amount that the wave's path is bent depends on the frequency, lower frequencies being diffracted more than higher ones. Thus a conversation overheard round the corner of an open door, appears in mumbled, low tones. Similarly low noises, like drum beats, can be better heard around buildings than high noises like whistles; this is why a distant band often seems to consist only of drums.

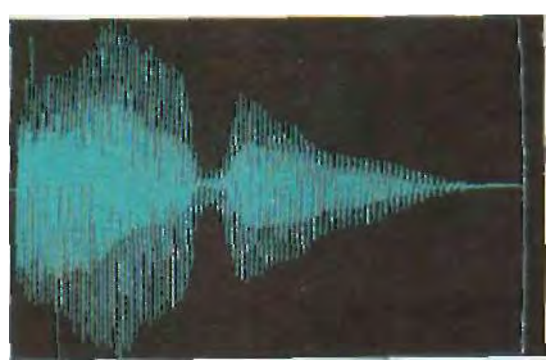
T *Two waves of the same frequency can cancel or reinforce, depending on their relative phase — the matching between peaks and troughs. Waves of different frequency (below) add together to give a complex waveform of varying amplitude.*

A *Reflection, interference and diffraction can be seen in this aerial photograph of waves in the sea. As the waves pass through a narrow gap, they spread out (diffract), and the interference of two waveforms is manifested in cross-patterned areas.*

Reinforcement



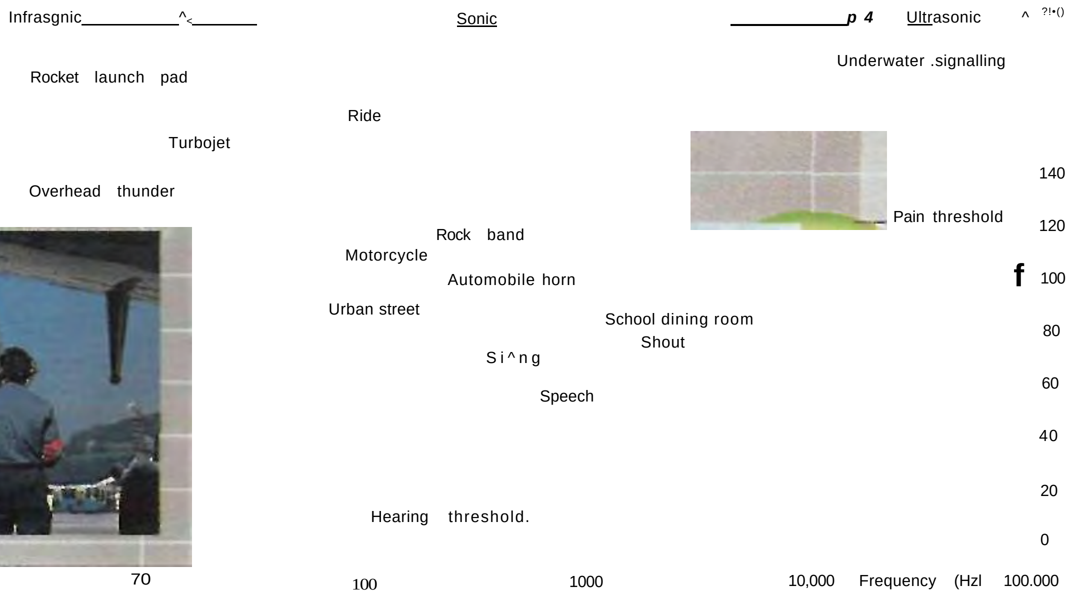
Complex_ wave



• *The waveform of "baby", spoken by a speech synthesizer.*

• 1 The human ear hears only a range of frequencies, being most sensitive to those around 5,000Hz. Sound levels above about 120dB relative to a zero dB level of W~"W/m'are painful, so the ears of people working close to jet engines, for example, must be protected /below).

The sound spectrum



1 70

Measuring loudness

The human ear perceives a sound wave of twice the intensity of another as rather less than twice as loud. Moreover, the ear responds to such a large variation in intensity that it is useful to define a scale that somehow compresses this huge range. The scale used is the "sound intensity level" scale. Its basic unit is the "bel", named after the Scottish-American inventor Alexander Graham Bell (1847-1922). However, the "decibel" (dB) - one-tenth of a bel - is more convenient to use.

The scale's zero point is defined as the threshold of hearing, at an intensity of 10 ^{12}watts/sq m, Other sounds are normally measured relative to this level. The scale is logarithmic, to approximate to the actual response of the human ear. Thus a WdB sound is 10 times as intense as one ofOdB, while a 20dB sound is 100 times, and 30dB 1000 times, as intense as the OdB sound.

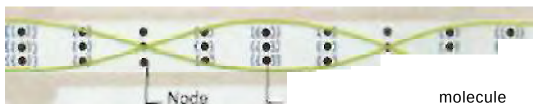
Pipes and strings

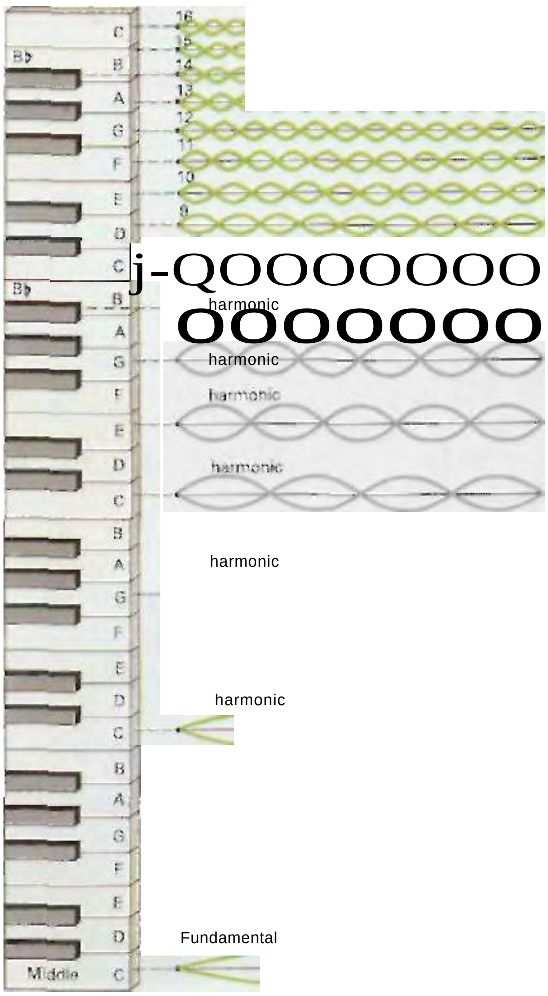
Most musical instruments produce sounds by setting a string vibrating or by initiating vibrations in a column of air. The basic process is to make the string or the air column vibrate at its own natural frequency, in other words to "resonate".

A stretched string, fixed at both ends and plucked at the center, will vibrate, the whole string moving from one side of its resting position to the other and back again. The vibration has a characteristic frequency which depends on the tension in the string, its weight and length. The shorter the string, the higher the frequency. The vibrating string sets the surrounding air molecules oscillating, generating a sound wave of the same frequency.

In a wind instrument such as a flute the musician sets air enclosed in a pipe in vibration. Air passes over a reed at the entrance to the pipe which causes eddies that generate vibrations in the column of air in the pipe. The frequency of the note produced depends on the length of the pipe, and whether it is closed at one end. The characteristic sound or "timbre" depends on the overtones that occur.

D2

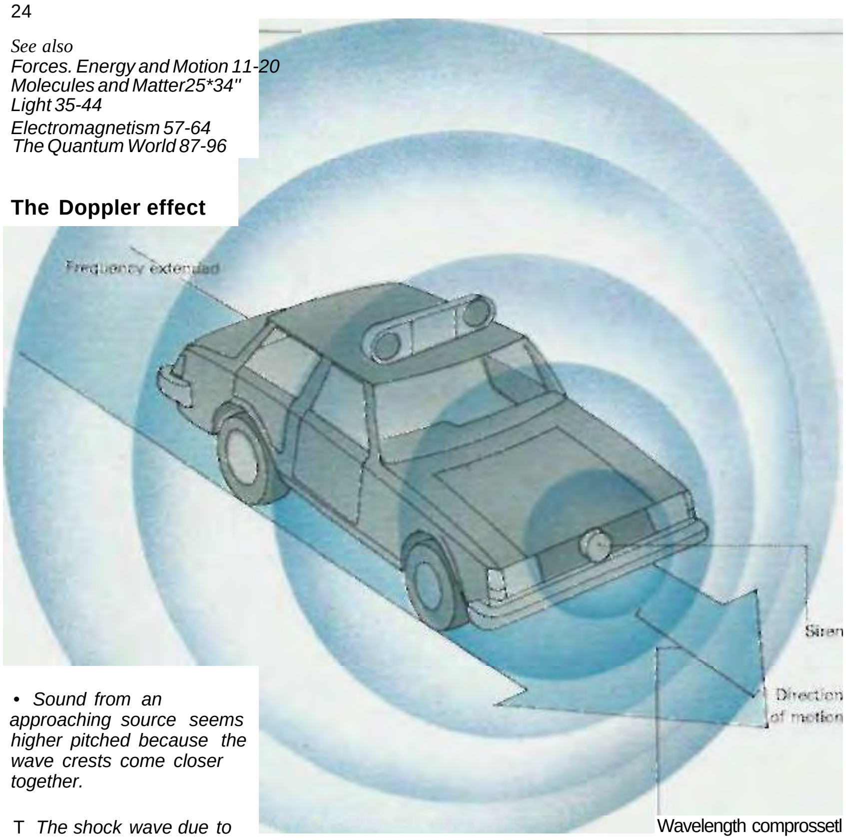




k -4 When a musician blows Into a wind instrument such as a recorder (above), the air vibrates, setting up a "standing wave" in the pipe. This wave does not move along the tube, but consists of a stationary pattern of air moving by varying amounts. Positions where there is no movement are called nodes, while movement is greatest at the antinodes, for example at the ends of the pipe. In the simplest standing wave, one wavelength fits within the tube; this corresponds to the fundamental frequency of this note. Notes of higher fundamental frequency are made by shortening the tube - removing fingers covering holes along the tube. But each note contains overtones. These are weaker waves of higher frequency which also have antinodes at the open ends. Similar standing waves are set up when strings are plucked or struck, as in a piano (left). Here in the fundamental mode the ends of the string are held fixed, while the center vibrates. The profile of the vibrating string maps out half a wave pattern. The keyboard shows how notes of higher frequency correspond to the overtones, or harmonics, of the fundamental middle C.

See also
Forces. Energy and Motion 11-20
Molecules and Matter 25-34
Light 35-44
Electromagnetism 57-64
The Quantum World 87-96

The Doppler effect



• *Sound from an approaching source seems higher pitched because the wave crests come closer together.*

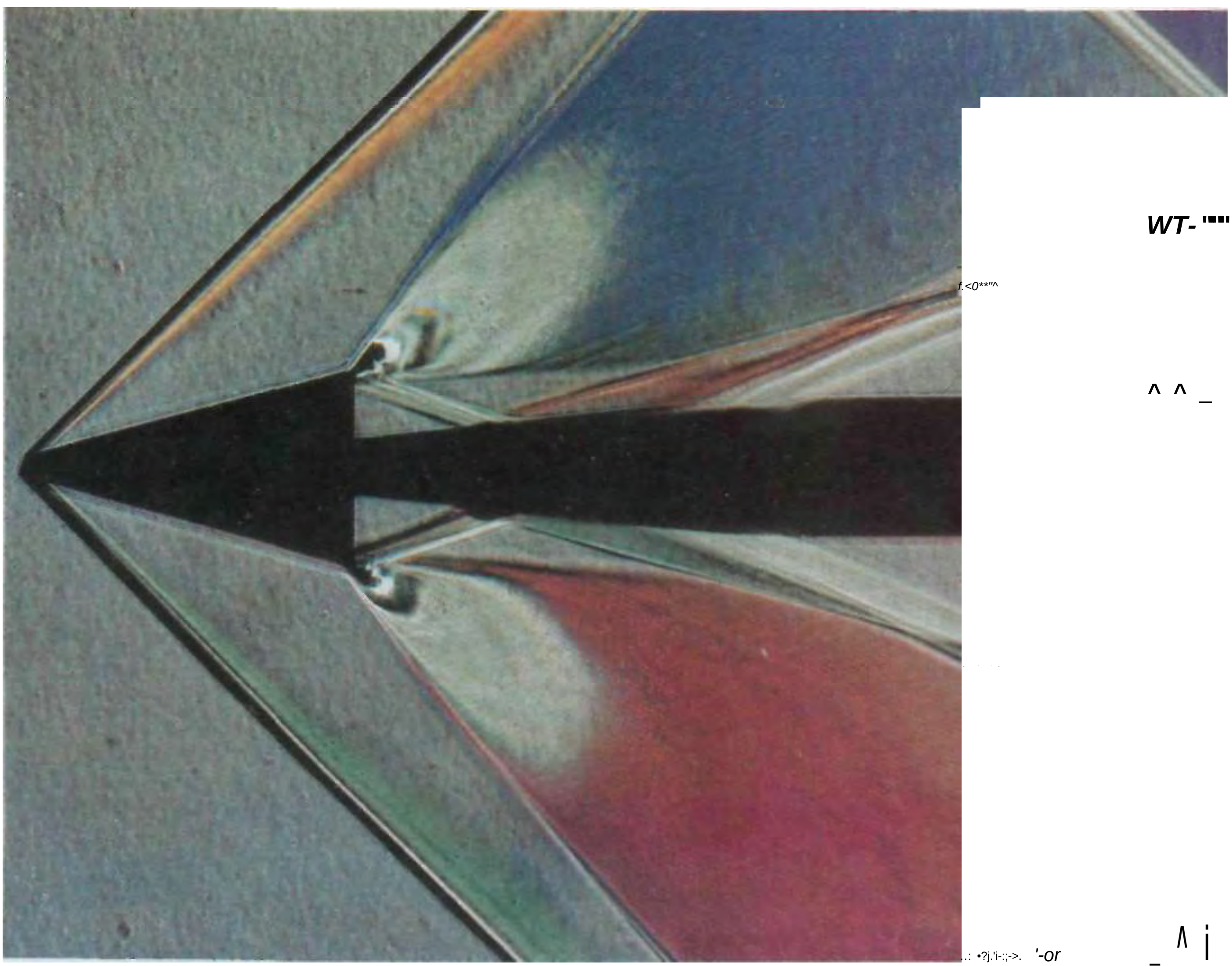
▼ *The shock wave due to a supersonic dart.*

The Doppler effect

A familiar wave phenomenon of sound is the change in pitch of the noise from a passing siren. This is an example of the Doppler effect, also observed for light waves. As the source of the sound moves closer to the listener, each successive compression is emitted closer to the previous one. The wave arriving at the listener is thus itself gradually squeezed together, so that its frequency appears higher as the source approaches. As soon as the source has passed, successive compressions are emitted at increasing intervals as the source moves away. The pitch of the sound drops.

Sonic booms

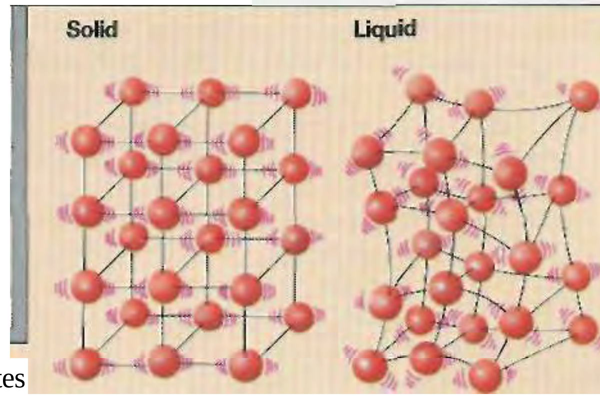
Sometimes the source of a sound travels faster than the waves it produces. A familiar example is the supersonic jet aircraft, which travels faster than the velocity of sound in the atmosphere. In such cases, the successive compressions arrive at the listener almost at the same time, and add together to produce a very loud noise. This "sonic boom" thus occurs continuously, and moves in the wake of the moving sound source, providing the source is moving faster than sound.



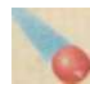
Molecules and Matter

4

Liquids, solids and gases...Oscillating molecules... Forces between molecules...Latent heat...Melting and boiling points... Viscosity... Thermodynamics... PERSPECTIVE...Pressure...Surface tension...Brownian motion.. .Stress and strain...Boltzmann...Boyle and the expansion of gases...Phase diagrams...The critical point..Amorphous solids and liquid crystals



Gas



A *In a solid, the attractive forces hold the molecules in a fixed framework although the molecules vibrate about their positions due to thermal energy. In a liquid, the attraction is weaker and the molecules can move around although they remain bound together. In a gas, thermal energy wins out over the attractive forces, and the molecules are free to move individually, spreading through large volumes.*

The matter of the everyday world exists in one of three familiar states or "phases" - solid, liquid, or gaseous. Solids have a fixed shape, are usually rather dense, and are very difficult to compress. Liquids are also rather dense and difficult to compress, but they differ from solids in having no fixed shape and are able to flow with varying degrees of difficulty. Gases usually have much smaller densities dian solids or liquids, are easily compressible, and flow even more easily than liquids.

A characterisitic feature of solids is that they often occur as crystals. Ice and gemstones are familiar examples of crystals, while modern electronics depend crucially upon crystalline silicon (| page 78). X-rays reveal that crystals are composed of a regular three-dimensional array of atoms spaced apart by a few tenths of a nanometer. These atoms are bound in place, but vibrate; these vibrations grow by increasing amounts as the crystal is heated. In gases, by contrast, the molecules are not fixed in position. They move about randomly in space with speeds that increase as the gas is heated.

Liquids also show some regular structure, but only across a few molecules and over very short intervals of time. The key difference between liquids and solids is that in a liquid some molecules are missing from their places. This leaves empty spaces into which other molecules may jump every so often. Most of the energy of the molecule goes in vibrating about a fixed position, as in a solid, but it is the movement of molecules from one place to another within the body of the liquid that gives it its properties of flow and viscosity.

Pressure

When a force acts upon an object, its effect depends on both how the force is distributed and what the substance is made of. For example, snow shoes spread a person's weight over a large area, so the wearer does not sink so easily into soft snow. But if the person wears shoes with spike heels, much of the same force is now concentrated into the small area of the heels, which now sink easily into grass. The difference lies in the pressure, which is defined as the component of force perpendicular to the area divided by the size of the area. So the same force exerts a larger pressure over a smaller area, and vice versa.

The effect of pressure on a material depends on the microscopic structure of the substance. Increasing the pressure squashes the molecules closer to each other. In a solid, the rigid structure means that very little change in volume occurs and the pressure is transmitted through the structure. In a liquid, the molecules move more freely, so the pressure acts in all directions as the molecules push against each other. That is why water will shoot out sideways through a hole in the bottom of a tank although the weight of the water is acting downwards. The same is true of a gas, but in this case the molecules are so far apart that an increase in pressure causes a decrease in volume as the molecules are squashed closer together.

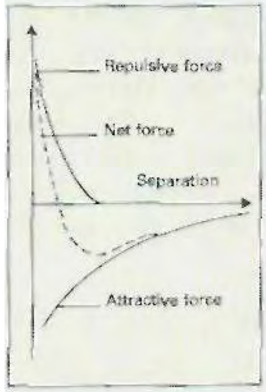
< *Experiments on air pressure became possible in the 17th century after the invention of the air pump by Otto von Guericke (1602-1686), a mayor of Magdeburg in Germany. Von Guericke himself performed a famous experiment demonstrating the pressure of the atmosphere, in which he showed how difficult it was to pull two hemispheres apart once the air had been pumped out from within them.*

Some materials, such as concrete, are able to resist compressive forces, but are very weak under tensile stress

Forces are required to hold molecules together. The fact that a single substance can exist in a solid, liquid or gaseous state reveals something about these forces. They must attract and repel other molecules, and be of short range. Without attractive forces the molecules would not coalesce to form liquids or solids; everything would be gaseous. Without repulsive forces matter would shrink to an infinitely dense point. The forces must decrease rapidly with increasing separation between molecules because physicists are able to describe the behavior of gases like air in everyday situations without reference to these forces; it is as if the molecules bounce apart from each other like billiard balls, even though they are separated on average by only a few nanometers. However, the attractive force must always be of longer range than the repulsive force if the molecules are to coalesce into a solid or liquid.

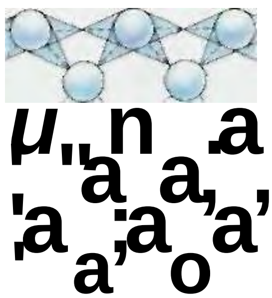
The force between a pair of electrically-neutral molecules, such as nitrogen, helium or water, decreases so rapidly with separation that only the "binding energy" between adjacent molecules is significant. Many properties of solids and liquids depend on the intermolecular forces and binding energy, and therefore many diverse physical phenomena are related to each other. The melting temperature, critical temperature, latent heat and surface tension are a few such related properties. The total binding energy of an assembly of molecules in the solid or liquid state is equal to the number of pairs of nearest neighbors, multiplied by the binding energy of a pair of molecules at their equilibrium spacing. At very low temperatures, this total binding energy is equal to the "latent heat of sublimation" - the energy needed to dissociate the solid into its separate molecules.

Some mechanical properties are also directly related to the intermolecular binding energy. The "elastic moduli" measure how hard it is to change the separation between molecules in a material by stretching, twisting or compressing it. Thermal expansion occurs because the attractive force is of longer range than the repulsive force. At a temperature above absolute zero the molecules vibrate about their lattice positions, and as the temperature is raised, the average separation between the molecules increases. The length of a piece of the material in bulk is governed by the average separation between molecules, so as the temperature rises, the material expands.



▲ The balance between a repulsive force and an attractive force – to give a net result of zero – gives the average separation between atoms and molecules in gas or liquid.

IP- As temperature rises, the average separation between atoms and molecules increases, causing thermal expansion. In bridges, this is allowed for by expansion joints.



•4 A Molecules at the surface of a liquid feel a net force pulling inward. This is surface tension. It provides a cohesive force between the surface molecules, which is sufficient to prevent the legs of a ripple bug from breaking through (left). The high surface tension in water is vital to many physiological processes.

Surface tension

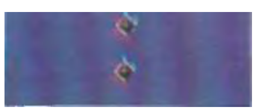
Within a liquid, the attractive forces between molecules pull in all directions, so the net effect on a single molecule is zero. But at the surface there is an imbalance. A molecule there is pulled more towards the body of the liquid than in the opposite direction. This effect is known as surface tension. In a drop of liquid, the intermolecular forces are tending to pull the surface towards the center. The result is a spherical drop.

*Surface tension can make a liquid climb "uphill" as when water climbs up a fine glass "capillary" tube. This happens because the attractive forces between the glass molecules and the water molecules are greater than those between the water molecules themselves. The surface of **the** water is pulled upwards, more so at the edges of the tube, creating a concave "meniscus". In other cases, such as with glass tubes and mercury, a convex meniscus forms and the liquid drops down the tube. This is because the forces between **the** glass molecules and the mercury molecules **are** weaker than those between the mercury molecules.*

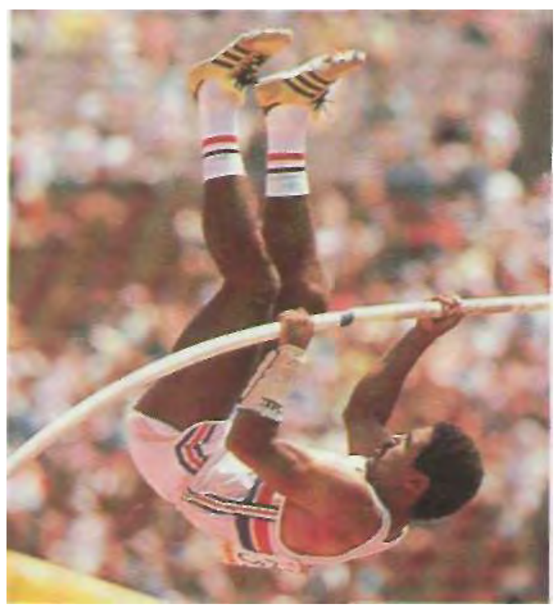
Young's modulus

If a length of copper wire is suspended from a support and a weight hung from the end, the force acting on the wire increases its length slightly. Provided the weight is not too great, the wire returns to its original length when the weight is removed. This is elastic behavior, and it is characterized by the modulus of elasticity, or "Young's modulus", in this example of a wire under tensile, or stretching, stress. Young's modulus is equal to the tensile stress (force per unit of area) divided by the change in length (also called the tensile "strain"), and its value for a material depends on the strength of intermodular forces.

There is, however, a limit to this elastic behavior beyond which permanent stretching of the wire occurs when the load is removed. The wire in this case stretches irreversibly, as layers of atoms slide permanently over each other. Quite large increases in length are possible before a "ductile" material like copper finally breaks. A "brittle" material such as glass will fracture almost immediately after the elastic stage has been passed.



A *Hardness of materials depends on the forces between their atoms and molecules. Here diamonds are fired at the surface of a metal to test its hardness.*



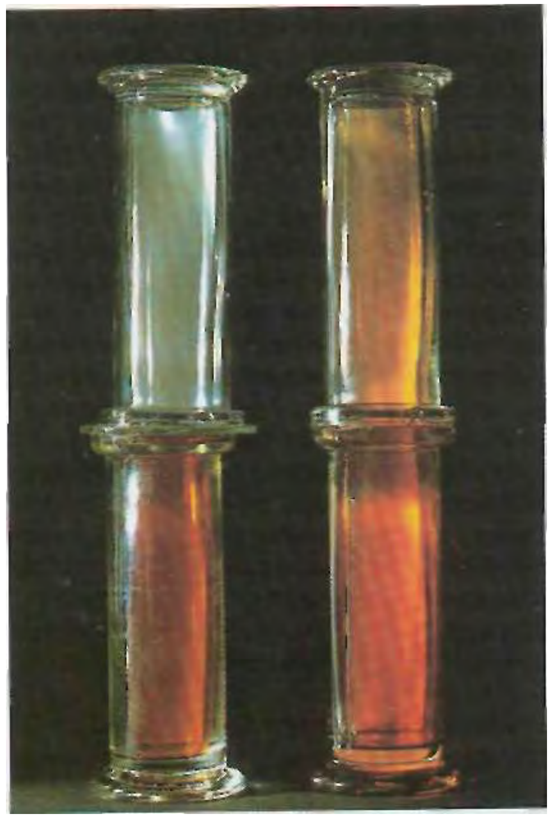
• *A pole vaulter uses the elastic properties of the pole to help gain height, as the bent pole springs back and flings the athlete through the air.*

The founder of "statistical mechanics ", Ludwig Boltzmann, committed suicide, depressed by the failure of others to appreciate his work

The temperature of a substance - its degree of "holness" - reflects the energy of the molecules it contains. The higher the temperature, the greater the *average* energy of the molecules. Not every molecule has identical energy - at any particular temperature, a range of energies is possible although not equally likely. The exact distribution of energy among the molecules depends on the temperature, according to a law due to the Austrian physicist Ludwig Boltzmann (1844-1906). Boltzmann developed a "statistical" theory for the behavior of matter, based on the average motions of the many atoms and molecules within a substance. "Boltzmann's law" refers to a system of particles in thermal equilibrium, and it states how the average number of particles with a certain energy varies with absolute temperature, and rises exponentially with rising temperature.

Boltzmann's law underlies many features in the behavior of materials at varying temperatures. Many chemical reactions, in both inorganic and biological systems, proceed much more rapidly as the temperature increases. For example, a change of one or two degrees in the processing temperature leads to a large change in the time required to develop a film. The reason is related to Boltzmann's law. The probability for a molecule to have sufficient energy to react chemically with another molecule increases very rapidly with temperature.

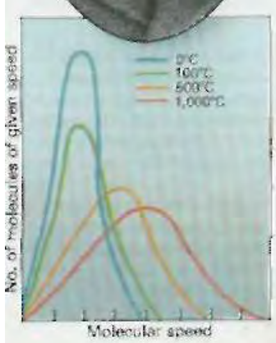
When a crystal is heated, the most energetic molecules break free from their positions in the lattice and migrate through the crystal. Other molecules move into the "holes" left behind. This phenomenon of "diffusion" is of major importance to the electronics industry in the manufacture of large-scale integrated circuits {% page 55). As the temperature rises, the number of holes in the crystal increases exponentially, according to Boltzmann's law, but while the holes are relatively far apart the substance still behaves as a crystalline solid. However, when the number of holes becomes very large there is a



A In a gas at room temperature, the molecules are moving around with a speed of nearly 500 m/s. This movement gives rise to diffusion, as the molecules spread out to fill any volume the gas enters. In this classic demonstration of molecular diffusion, bromine, the brown gas, and air {left) can be seen to mix once a plate keeping them apart has been removed (right).

Ludwig Boltzmann

Boltzmann was born in Vienna in 1844. This was the era in which the theory of thermodynamics began to emerge (} page 34). At the same time, the kinetic theory of gases was also being developed showing how properties such as pressure could be understood in terms of the overall behavior of many atoms. Boltzmann's great achievement was to discover the links between these two apparently different theories, combining the thermodynamic properties of bulk matter and the microscopic world of kinetic theory. Using a statistical treatment of the average mechanical behavior of individual atoms, he deduced the thermodynamic properties and founded the theory of "statistical mechanics". His work bridged the classical theories of the 19th century and the quantum theories ($ page 87) of the 20th. Yet this was at a time when atoms were not accepted by all scientists. He committed suicide at the age of 62, depressed by the failure of his fellows to appreciate his work. His tombstone is inscribed with the equation that encapsulated his statistical interpretation of "entropy" (} page 34).



• Boltzmann was the father of statistical mechanics, which is used to study the average behavior of large collections of atoms. His work was based on foundations laid in particular by the Scottish physicist James Clerk Maxwell 11831-1879), who first worked out the distribution of velocities for gases at different temperatures.

Brownian motion

Molecules are too small for their movement at high speeds to be seen directly. However, in 1827 the British botanist Robert Brown (1773-1858) first observed with a microscope the abrupt, random movements of very small solid particles (pollen grains) immersed in a liquid. These random jumps result from the impacts of molecules in the liquid on the particles, as required by Boltzmann's law.

The French physicist Jean Perrin (1870-1942) obtained further proof of Boltzmann's law early this century. He suspended in water microscopic particles of resin (having a density only slightly higher than water), and counted them at different heights using a microscope. The variation in their number with height was exactly what Boltzmann predicted. In the same experiment, he measured "Avagadro's number". Amedeo Avagadro (1776-1856), was an Italian physicist who first put forward the notion that equal volumes of gases, at the same temperature and pressure, contain the same number of molecules. Avagadro's number is the number of atoms in 12g of carbon-12, or 6.02×10^{23}.

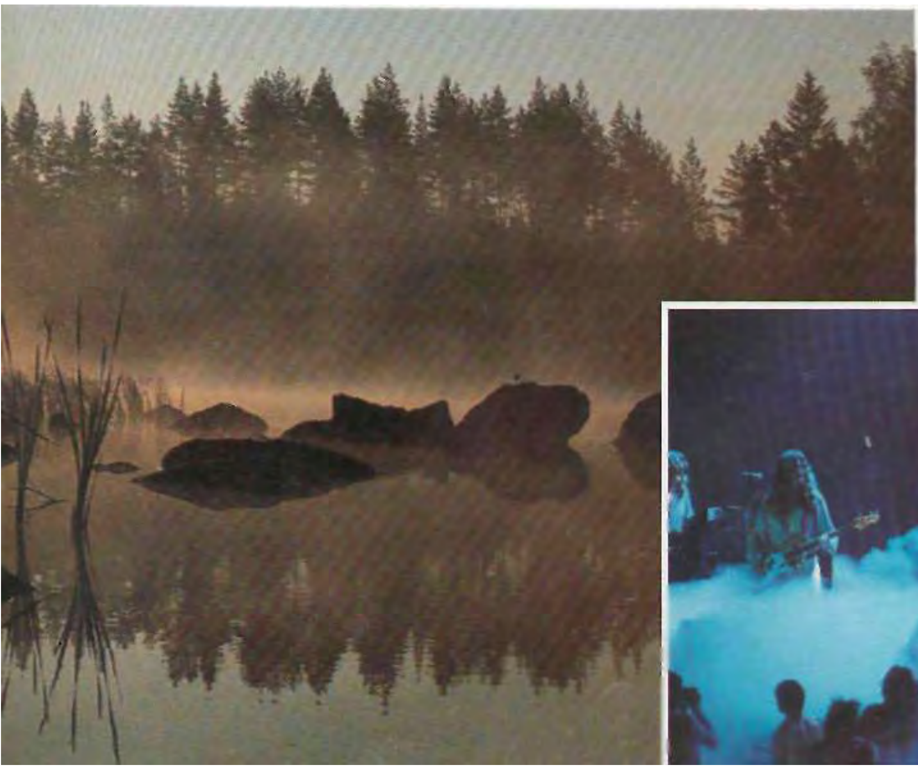
Measuring the Sun's temperature

The French astronomer Audouin Dollfus (b. 1924) used Boltzmann's law in 1953 to measure the Sun's corona. The Sun emits a red line from highly-ionized iron atoms. This line should be extremely narrow, but in the Sun's corona it is considerably broadened. Dollfus interpreted the broadening as due to wavelength shifting of the light emitted by molecules moving towards or away from an observer on Earth. He calculated the distribution of molecular speeds from these data, and found that it fitted well with the predictions of Boltzmann's law for a coronal temperature of 2.1 million degrees K.

good chance of adjacent lattice sites being vacant. The forces holding the nearby molecules in place are then greatly reduced and these molecules start to move about inside the crystal. This molecular mobility is manifest as "melting". As more heat energy is added to the melting solid it releases more molecules from their lattice sites. The temperature of the substance, meanwhile, remains constant until it is completely liquid. The heat energy required to melt a substance completely is called the "latent heal of melting".

Physicists can also explain the evaporation of a liquid in terms of the Boltzmann distribution of energies of the molecules. Much more energy is required for a molecule to break away from its neighbors and leave the liquid completely than for the molecule to change from one set of neighbors to another and move about in the liquid. At room temperature, for example, only a tiny fraction of water molecules have enough energy to evaporate. Nonetheless, a bowl of water will completely evaporate away over the course of a few days because the water slowly absorbs heat from its surroundings, and ultimately all the molecules will have acquired sufficient energy to escape. When a liquid is heated to higher temperature, however, a much larger fraction of molecules has the required energy, and it evaporates much faster.

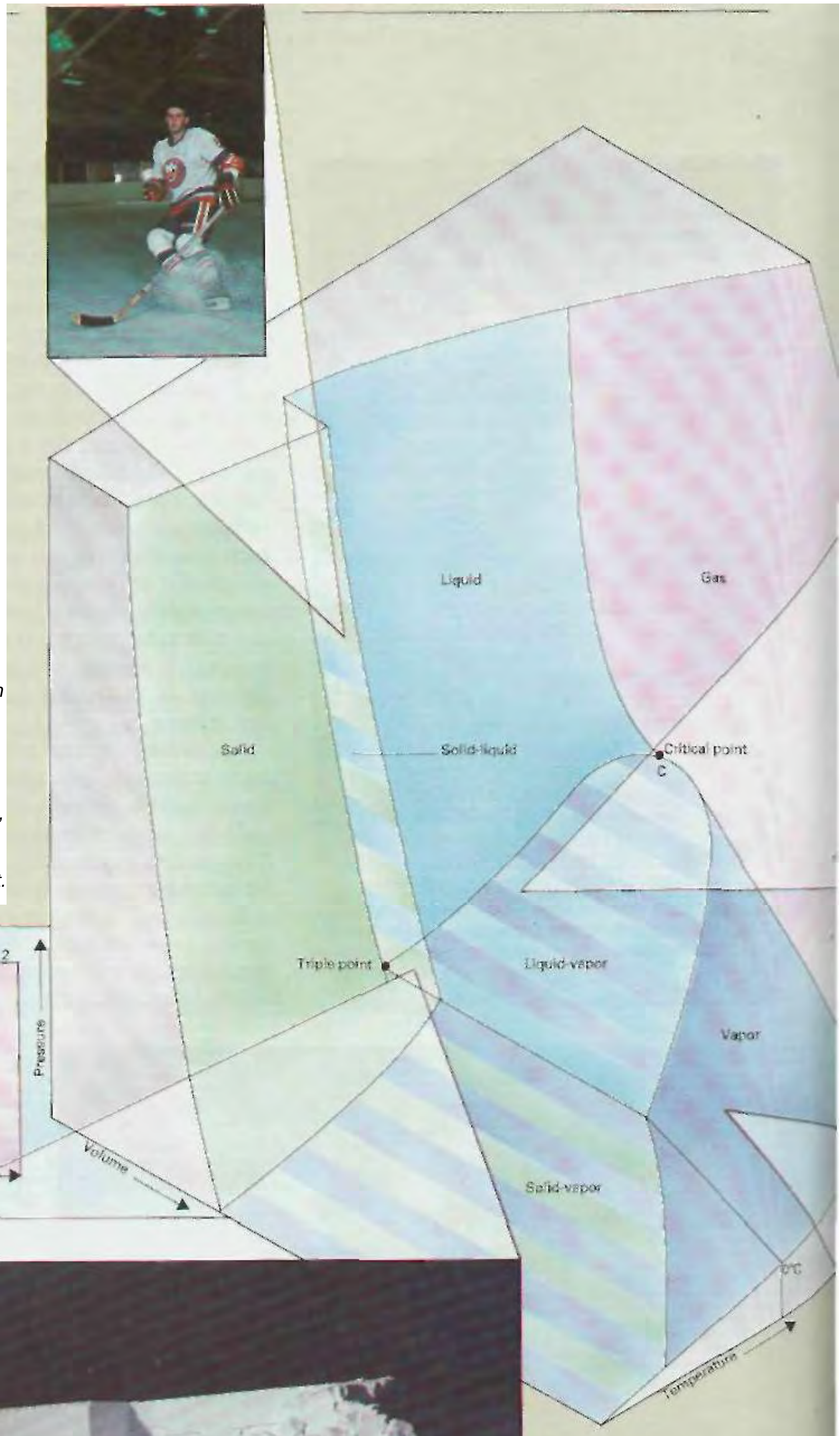
"Sublimation", the evaporation of a solid directly into a gas, is most commonly and spectacularly seen at the theater when Cardice (solid carbon dioxide refrigerated at low temperatures) is thrown onto the stage to produce clouds of vapor looking like mist or fog. The energy required for a molecule in a solid to break away from its neighbors and evaporate is even greater than for a molecule in a liquid, and so we are not usually aware of ice, for example, subliming to water vapor. Even so, washing hung outside in sub-zero temperatures will eventually dry, because some of the molecules still have enough energy to escape.



•< Mist gathers above a lake as the Sun rises in the early morning. Evaporation occurs when molecules in a liquid break away totally to form a gas. According to Boltzmann's law, the probability for a molecule to have enough energy to do this increases exponentially with temperature-puddles soon evaporate after a downpour on a hot day.

T Sublimation occurs when molecules have sufficient energy to escape directly from a solid to form a gas. This effect is seen in a theater, when solid carbon dioxide, or Cardice, which has been kept refrigerated, is thrown onto the stage to produce fog-like clouds of vapor as it sublimes. Sublimation requires more energy than evaporation.

Ice floats on water because water becomes less dense when it freezes

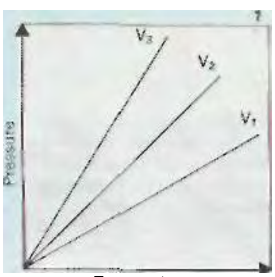


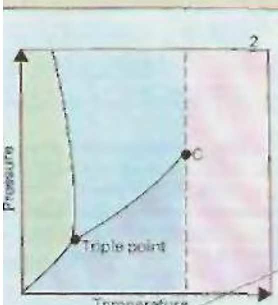
Pressure, volume and temperature

Whether a substance exists as a solid, a liquid or a gas depends not only on the temperature, but also on the pressure exerted on the substance and the volume it occupies. One of the first to study the relationship between these quantities was the Anglo-Irish physicist and chemist Robert Boyle (1627-1691). He showed that the product of pressure and volume for a fixed mass of gas at fixed temperature is approximately constant: double the pressure and the volume is halved.

Others extended this work by varying the temperature, and found that the pressure falls in proportion to decreases in temperature. Their results gave rise to the concept of an "absolute zero" of temperature, corresponding to a (hypothetical) zero pressure. Measurements indicated that this should occur at 273°C below the freezing point of water at atmospheric pressure. Thus -273"C became the starting point of the "absolute" scale of temperature, which has the same size of degree as the Celsius scale. Temperatures on this scale are referred to in terms of degrees (K), after the British physicist Lord Kelvin (1824-1907) who did much important work on the theory of heat and temperature.

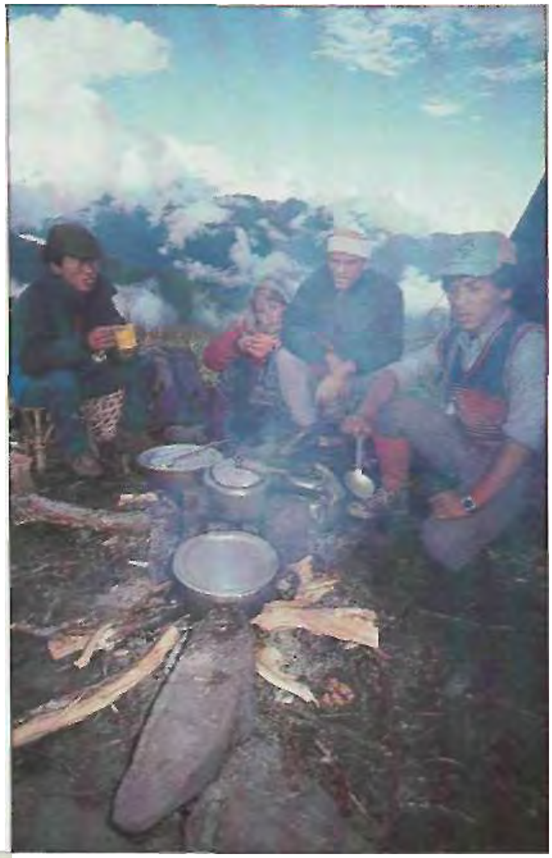
The early work on gases by Boyle and others can be summarized in a single relationship: pressure times volume equals a constant times temperature, where the temperature is measured on the absolute scale. Both the "gas constant" and the volume are proportional to the mass of gas present.





A • *The "pressure law" relates the pressure of an ideal gas (with negligible intermolecular forces) to its temperature for fixed volumes, V(1). If extended to low temperatures the lines for different volumes all meet at zero pressure and the absolute zero of temperature. A real gas 12) changes to liquid and solid phases, however, as the temperature falls and the force between molecules is no longer negligible compared with their kinetic energy. In water, shown here, increasing pressure can revert solid (ice) to liquid, and this partly explains the slipperiness of ice. Water becomes less dense on freezing, so ice floats on water.*

•* The relationship between the phases of matter-solid; liquid, gaseous -at the different variables of temperature, pressure and volume, can be plotted on a single, three-dimensional "phasediagram". The diagram for water is shown here. The boundaries between the phases are plotted. Water is unusual in expanding when it freezes, and this is shown on the diagram by the notch in the face between solid and liquid. The point C is the critical point, the highest temperature at which the substance can exist as a liquid. It represents the substance's highest boiling point; below this, the boiling point varies with pressure. A gas below the critical point is known as a vapor.

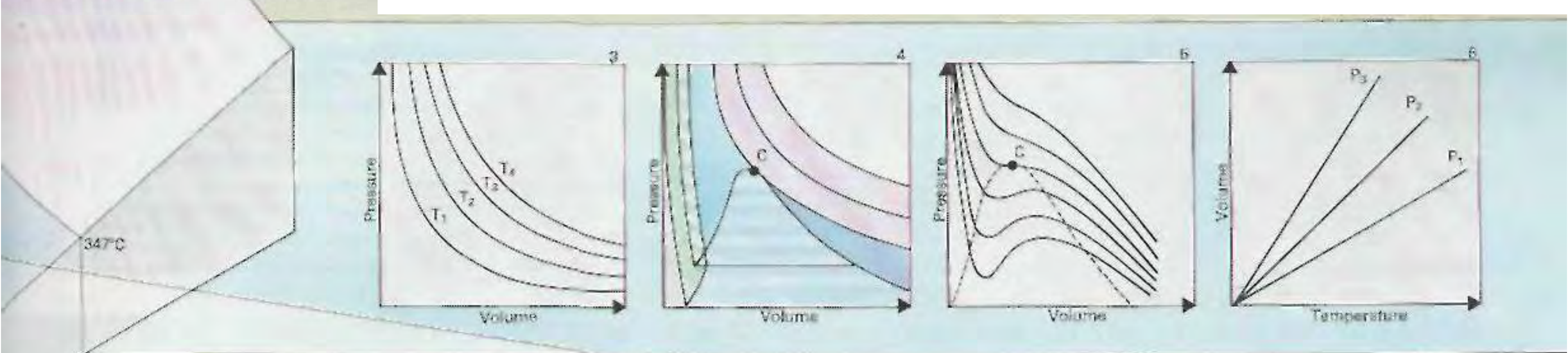


Phase diagrams

The values of pressure, volume and temperature related by the "equation of state" described opposite lie on a curved surface in a three-dimensional space depicted in a "phase diagram ". Such a diagram is based on three axes which represent pressure, volume and temperature.

The surfaces corresponding to the simple equation of state occur on the phase diagram only in the region where the substance behaves as an "ideal" gas. Elsewhere different relationships hold between pressure, volume and temperature. Only for certain ranges of these quantities can a substance exist in a particular phase such as a solid, liquid or gas. Over other ranges, two phases, such as solid and liquid, coexist in equilibrium; and for one particular value of temperature and pressure all three phases are in equilibrium. This condition is shown as the horizontal line where the liquid-gaseous and the solid-gaseous phase boundaries meet.

Adding heat energy causes some of the solid to melt and some of the liquid to evaporate, so that the volume increases, but both the temperature and pressure remain constant. These values of temperature and pressure define the "triple point", which for water corresponds to a temperature of -0.1"Cata pressure of somewhat less than one hundredth of atmospheric. Ice and water are in thermal equilibrium at atmospheric pressure at a slightly higher temperature - the melting point of ice.



A • The pressure and volume of an ideal gas at constant temperature IT) are related by the smooth curves of Boyle's law (3). But this applies to a real substance such as water only at high temperatures 141. A better description of real gases comes from Van der Waals'theory, which takes into account the forces between molecules (5), although this does not describe the transitions to liquid and solid phases. In real substances, the temperature for boiling varies with pressure; so tea brewed at low pressure on a mountain boils at lower temperatures. Pressure changes in the fluid flow around a propeller can make bubbles of gas form.

The behavior of real gases

Two hundred years after Boyle, the Irish physicist Thomas Andrews (1813-1885) made extensive measurements on carbon dioxide and drew up phase diagrams which reveal the difference in behavior of carbon dioxide from that of an ideal gas at high pressure and low temperature. These differences led the Dutch physicist Johannes van der Waals 11837-1923) to describe how real gases behave. He argued that the molecules of a gas take up space, so that the equivalent volume of an ideal gas is a little smaller than the measured volume of a real gas. And the measured pressure of a gas is smaller than the ideal gas pressure because of the net attraction between the molecules. Van der Waals' equation describes the behavior of real gases well over quite a wide range of pressures and temperatures. It fails only at high pressures and low temperatures where the separation of the molecules is similar to that in a liquid or solid.

Many different physical phenomena can be explained in terms of forces between molecules in a substance

Explaining the properties of matter

The molecules in a gas move about with very high speeds. At room temperature and atmospheric pressure, for example, the average molecular speed of the nitrogen and oxygen molecules in air is about 450 meters per second. However, the average distance traveled by a molecule before it collides with another molecule is very small (less than one ten-millionth of a meter). The diffusion of molecules from one region to another therefore involves many millions of molecular collisions. This process of diffusion explains the thermal conductivity of gases and also their viscosity. In the absence of convection currents or radiation, heat is transferred from a hot region of a gas to a cooler region by molecular collisions. Molecules in the hot region travel faster, and when they collide they give up some of their excess energy, thereby heating up the cooler regions of gas. Viscous forces can also be understood in terms of molecular collisions and diffusion.

Theory predicts that thermal conductivity and viscosity of gases should not depend on pressure, but should increase with the square root of the absolute temperature. By contrast, the viscosity of a liquid decreases with temperature, indicating that the mechanism of diffusion in a liquid is quite different from that in a gas. Indeed, diffusion in liquids is very similar to .that in solids, and occurs because molecules jump into adjacent, vacant lattice sites. At higher temperatures the number of such holes increases strongly and so does the rate of diffusion. If it is easier for a molecule to move in one direction than another there will be a net rate of diffusion in this direction.

The viscosity of a liquid increases rapidly with pressure, in contrast to the behavior of a gas. This is because the effect of the pressure is to squeeze the holes and make it much more difficult for a molecule to force its way into an adjacent vacant position. This effect is of great importance in engineering. Many sliding mechanisms operate successfully only because the lubricating oil is not squeezed out.'Heavily loaded gear teeth may enmesh with a contact pressure of several tonnes per square centimeter, at which pressure the viscosity of a typical lubricant may have increased a million-fold.

Although the description of a substance as being in the solid, liquid or gaseous phase is convenient, it can be misleading for it applies strictly to the properties of ideal substances. Such common substances as ice, pitch and lead flow like very viscous liquids when large forces and pressures are applied to them; water shows rigidity, a property of solids, if one attempts to change its shape too rapidly; gases moving in bulk close to the speed of sound can sustain sharp changes in density and pressure over quite small distances ... and alloys, plastics and glasses are all much more complicated to classify.

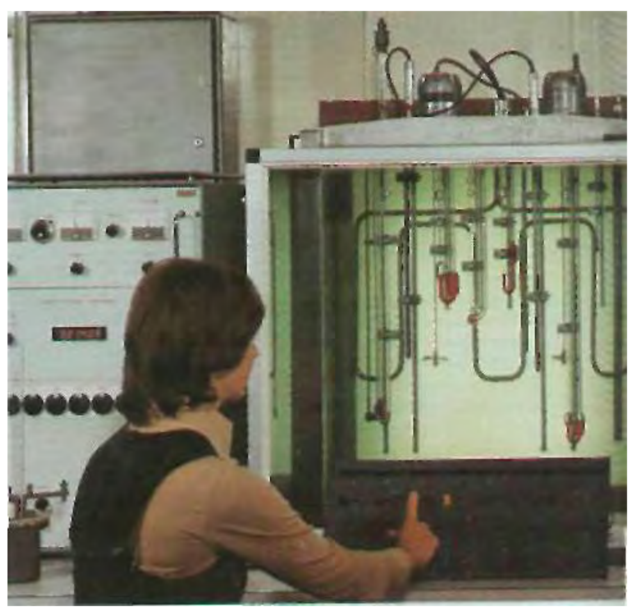
A simple understanding of matter is possible only under the rather special conditions of gases at high temperatures and low pressures (when intermolecular forces are negligible) or crystalline solids at very low temperature (when imperfections and diffusion can be ignored). The liquid state, in particular, is difficult to understand. Physicists can describe some features, such as evaporation and superheating, by thinking of the liquid as a very dense gas. Other features, such as the tensile strength and viscosity of a liquid, can be understood only by considering the liquid as an imperfect solid. There is, however, a link between these two, which is apparent in the behavior of a substance at pressures and temperatures where separation between molecules remains close to that of the liquid phase.

Amorphous solids and liquid crystals

Most materials are naturally crystalline in their solid form. Some materials have no regular structure; they are "amorphous". In these solids, the atoms do not form in a regular pattern.

Some amorphous materials, such as rubber, consist of molecules in the form of long chains (polymers) which have become tangled together. In other instances, the solid is like a "supercooled" liquid in which the irregular pattern of atoms of the liquid state has become "frozen in". Glass is perhaps the most ubiquitous amorphous solid. It is made from a mixture of soda and lime with sand, all of which fuse together in a liquid at temperatures of around 1500°C. The glassy state forms as the liquid cools and rapidly becomes viscous, preventing crystals from forming as it solidifies.

Liquid crystals are, by contrast, liquids with a structure like a solid. They are liquids in which a high degree of ordering can occur, for example when an applied electric field organizes the normally random arrangement of the molecules. This can alter the optical properties of the liquid crystal, as in the displays on electronic watches.



• *Ice in a glacier flows slowly downhill, like a very viscous liquid, influenced by tremendous forces.*

•4 *Glass is perhaps the most familiar amorphous solid, used for centuries in windows for example.* ***Its*** *irregular atomic structure, like that of a "frozen" liquid, characterizes other "glassy" materials.*

T *An array of rod-shaped liquid crystals seen in polarized light, showing one of the several possible regular arrangements of the crystals.*

f Measurements of the viscosity of liquids such as lubricating oils are crucial to industry. Here, instruments for measuring viscosity are being calibrated.

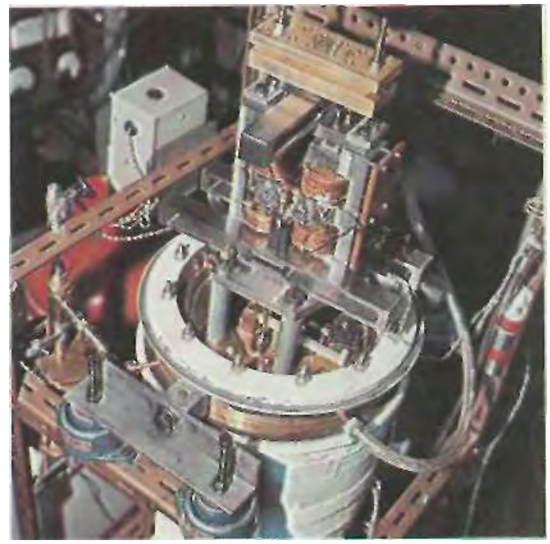
See also
Forces, Energy and Motion 11-20
Sound 21-4
Light 35-44
The Quantum World 87-96

Thermodynamics

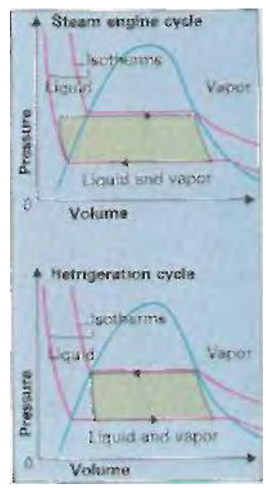
The pressure of a gas, the viscosity of a liquid and even the rate of a chemical reaction can change with temperature. Although these changes occur in widely differing systems, there is a generalized framework that can be applied to them. This is thermodynamics.

"Classical" thermodynamics was developed around 1850 by the Scottish physicist William Thomson (1824-1907) and the German Rudolf Clausius (1832-1888). These men built upon work by Frenchman Sadi Carnot (1796-1832), who in 1824 published a treatise on heat engines - engines that use heat to perform work. He proved that no engine could be more efficient than his idealized engine, operating a reversible cycle between two temperatures, and thai the efficiency depends on the temperatures between which the engine operates.

Carnot's insight became enshrined in two laws of thermodynamics. What is now known as the first law is a statement of the conservation of energy (| page 18), with heat taken into account. The first law showed that heat supplied to a system goes both in doing work and in changing the internal energy of the system. The second law is that heat cannot flow from a colder to a hotter body, without some other changes occurring. The second law reflects a basic lack of symmetry in the physical world: processes that can occur spontaneously in one direction will not occur equally well in reverse. If a partition between compartments containing two gases is removed, the gases will eventually mix; but they will not "unmix" again. The "arrow" that imposes this kind of direction is known as "entropy". Entropy, like energy, is a property of a system that changes when heat is supplied. In controlled conditions, the change in entropy is equal to the heat supplied divided by the temperature. Understanding of matter based on atoms leads to a deeper insight into entropy as a measure of "order" in a system. When the gases mix they become less ordered: entropy increases. And this reveals a more general statement of the second law - only those processes occur naturally in which entropy increases. Overall, energy is conserved (the first law) but entropy rises (the second law).

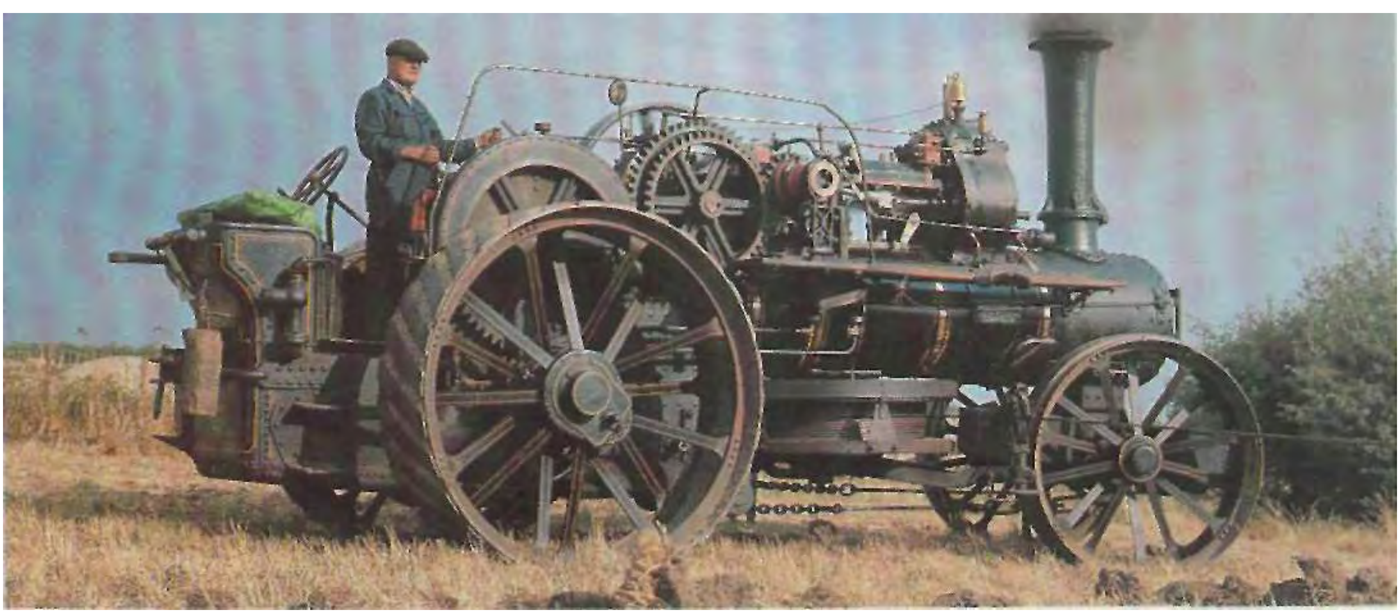


A *This methane-fueled generator is an example of a Stirling engine, the most efficient and cleanest form of engine yet devised.*



> *A steam engine works by allowing pressurized steam to expand and push a piston as the temperature falls. The resulting mix of liquid and steam is condensed to liquid and the pressure increased before reheating in the boiler. A refrigerator works on a similar cycle operating in reverse.*

T *The steam engine revolutionized work, from industry to agriculture, throughout the 19th century.*



Light

5

Light rays...Lenses and mirrors...Reflection and refraction...Prisms...Colors...The wave nature of light... Polarization...Light and particles...PERSPECTIVE... Measuring the speed of light... White light...Einstein and special relativity...Flying clocks around the Earth

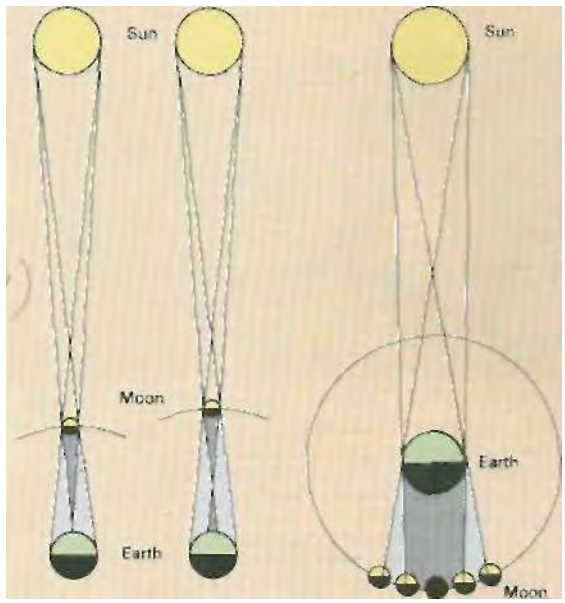
V A *A total eclipse of the Sun in 1980*

Lr~

Simple observations reveal some of the more obvious characteristics of light. The outlines of shadows in strong sunlight show that light travels in straight lines, at least on a macroscopic scale, and sunshine filtering into a room through small openings appears to form well-defined beams. This gives rise to the idea of "rays" of light, a concept that is useful in appreciating some of the basic properties of light, as well as the operation of many optical instruments. But what is light?

By the early 18th century, scientists had found that they could explain many optical effects in terms of the general properties of waves (f- page 48). As with sound waves (| page 22) and ripples on the surface of water, light can be seen to undergo reflection, refraction, interference and diffraction. However, what exactly constitutes a "light wave" was not answered until a century later, with the work of the British physicist James Clerk Maxwell (1831-1879). Maxwell drew together many observations concerning electricity and magnetism and incorporated them in a single theory which predicted the existence of "electromagnetic waves" O page 57). According to this theory, these waves travel through a vacuum at a velocity given by two constants related to electric and magnetic units, and this velocity is the same as the velocity of light. This was a revelation: it showed that light is an electromagnetic wave with a particular range of wavelengths. It forms part of a huge spectrum that ranges from gamma rays to radiowaves.

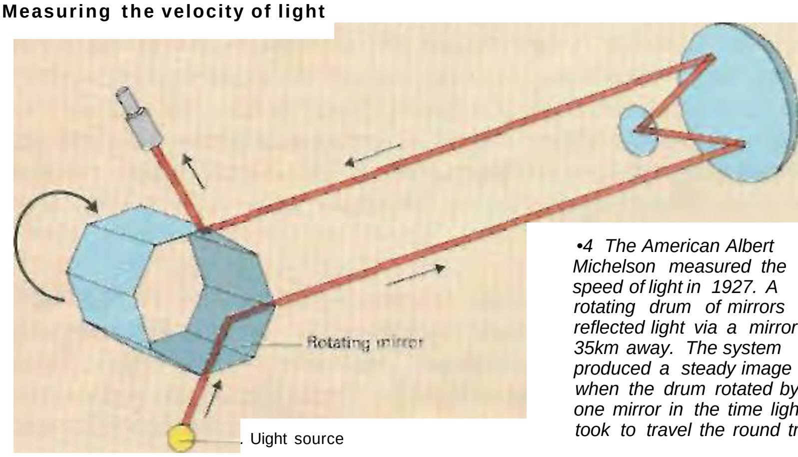
For many purposes the wave theory of light is adequate. But when it comes to explaining the absorption and emission of light on the atomic scale, the wave description is not tenable. Light must then be described as packets of energy, or "particles" of light, called photons, which interact individually with electrons in atoms. The discovery of the dual nature to light, demonstrating that it behaves both as waves and particles, brought about the development of quantum theory (• pages 87-96) and revolutionized physics in the 20th century.



3#€> (*<§>€ ©(*•€©

A Eclipses of the Sun Heft) and Moon (right) demonstrate how light travels in straight lines, casting shadows over great distances. On a much smaller scale, light can bend round corners, however, when it diffracts § page 38).

Measuring the velocity of light



. Uight source

•4 The American Albert Michelson measured the speed of light in 1927. A rotating drum of mirrors reflected light via a mirror 35km away. The system produced a steady image when the drum rotated by one mirror in the time light took to travel the round trip.

Measuring the velocity of light

The velocity of light in a vacuum, 299,792-5km/s, is one of the fundamental constants of physics, usually denoted by the letter c. *As Einstein showed in his special theory of relativity (+ pages 42-43), this is a universal "speed limit". Nothing can travel faster than the speed of light in free space.*

The Danish astronomer OlafRoemer (1644-1710) determined the velocity of light in 1676. Roemer measured the times at which one of the moons of Jupiter emerged from the shadow of the planet. He found that the end of such an eclipse occurred later when the Earth was further from Jupiter. A knowledge of the orbits then gave a value for the velocity of light. The result was lower than the correct value by about 25 percent, but it confirmed that light travels much faster than sound.

The refraction of light as it crosses the boundary between two substances depends on the relative speeds of light in the two materials

• T *Light rays "bend"- change direction - when they pass from one material to another. This is the process of refraction and it is related to the difference in the velocity of light in differing materials. It is put to use in lenses which can converge light rays (left), bringing light parallel to the axis of the lens together at a single point, the focus F. Lenses can also make parallel rays diverge (right), as if from a single point.*

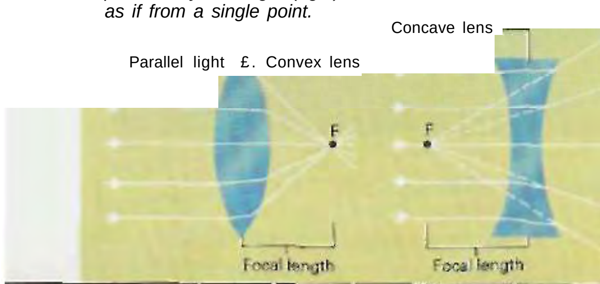
Angle of incidence

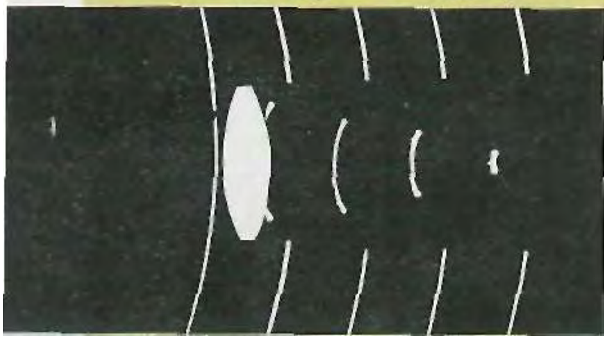
Glass

Angle of refraction

Concave lens

Parallel light £. Convex lens





• *A lighthouse uses both mirrors and lenses to produce a powerful, focused beam that can sweep across the horizon. A solid lens would need to be too big and heavy to be practical. Presnel lenses are used instead, in which the lens surface is divided into a series of concentric circles. Relatively thin sections of glass are set in each ring at the angles that would be found in a solid lens at that point.*

M *This image of the focusing power of a lens was produced by superimposing a sequence of high-speed holographic photographs of light pulses, using laser pulses of 10 picoseconds each. As well as showing how light is brought to a focus by the lens, the picture shows that the light is slowed as It passes through the glass — the focused pulses are delayed relative to the original beam.*

T *A converging (convex) lens produces an enlarged but inverted image of an object placed between the focus and a point at twice the focal length. This is a "real" image, which means that it would appear on a screen placed at its location, but cannot be seen by eye. A diverging (concave) lens, on the other hand, always produces an upright, diminished "virtual" image, which cannot be formed on a screen but which can be seen through the lens.*

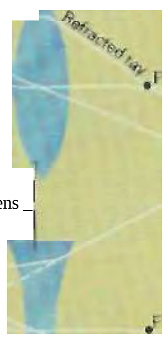
Incident ray

Object

Lens

Object

Virtual image



Real image

• *A wide beam of light is not all brought to a focus at the same point, because the angle of incidence varies toward the edge of the lens. This blurs the image - an effect known as spherical aberration.*

Parallel light

Reflection of light

The 17th century saw the invention of the first microscopes and telescopes, and the first theories of light. By this time scientists were aware of two laws governing the behavior of light, in addition to the fact that it seems to travel in a straight line. Both these laws concern what happens to light rays when they meet a surface, for example between air and glass.

The law of reflection states that the angle of reflection equals the angle of incidence (the angle at which a light ray strikes the surface), and that the reflected and incident rays both lie in the plane that contains a line at right angles to the surface. Simple diagrams using this fundamental law show how mirrors create images. With a flat mirror the eye sees light that appears to come from behind the mirror. In fact, the light has been reflected and the image seen is not a real image, but a virtual image: no rays connect the image to the observer's eye. A convex mirror also produces a virtual image, this time reduced in size. A concave mirror can produce an image between the eye and the reflecting surface; although inverted, this is a real image, because light rays do connect the eye and the image.

Refraction and Tenses

The second law concerns the "refraction" or bending of light as it crosses the boundary between two substances. This law states that the angle of refraction is in a constant relationship to the angle of incidence. The Dutch scientist Willebrord Snell (1591-1626) first enunciated this law in 1621, and it has since become known as Snell's law,

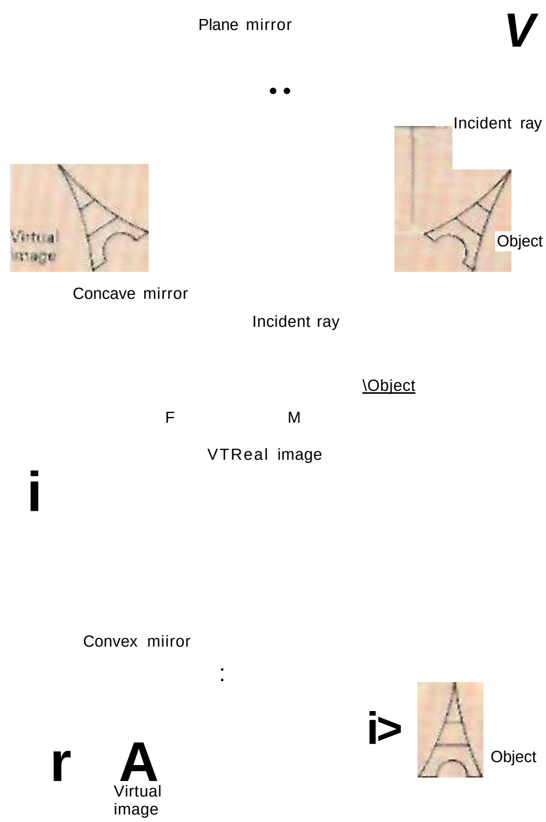
^ *Astronomical telescopes generally use curved mirrors. The first telescopes were based on lenses but the problem of chromatic aberration It page 38) led Newton to build the first reflecting telescope in 1671. Mirrors can be built with much larger diameters than lenses, and are the natural choice for large telescopes designed to collect as much light as possible. This mirror is for the Space Telescope.*

although in France it is known as Descartes' law after the French philosopher Rene Descartes (1596-1650), who rediscovered it some years later. The constant here depends on the nature of the substances on either side of the boundary. If the incident ray is in a vacuum, then the constant gives the "refractive index" of the refracting material. The angle of refraction is always smaller than the angle of incidence when a ray of light enters a denser medium and larger when it enters a less dense medium. The refractive index of a material is also equal to the velocity oflight in a vacuum divided by the velocity of light in that material. Thus glass, with a refractive index of about 1-5, slows light down to about 200,000km/s.

.SnelPs law explains why a pool of water appears to be shallower than it really is, and why an object such as a spoon seems to bend as it is lowered into water. In both cases the eye sees a virtual image, of the bottom of the pool or the lower half of the spoon, and this is displaced from the position of the actual object by the bending of the light rays. Refraction also underlies the operation of lenses.

A beam of parallel light rays that pass through a convex lens converges to a point on the other side of the lens. This point is called the "focus", and the distance between the center of the lens and the focus is the "focal length". (A concave mirror converges parallel light in a similar way.) A convex lens makes a simple magnifying glass if the observer holds the lens so that the object being viewed lies between the lens and its focus. In this case, the lens forms an enlarged virtual image of the object. A concave lens diverges parallel light, so that it appears to come from the focus.

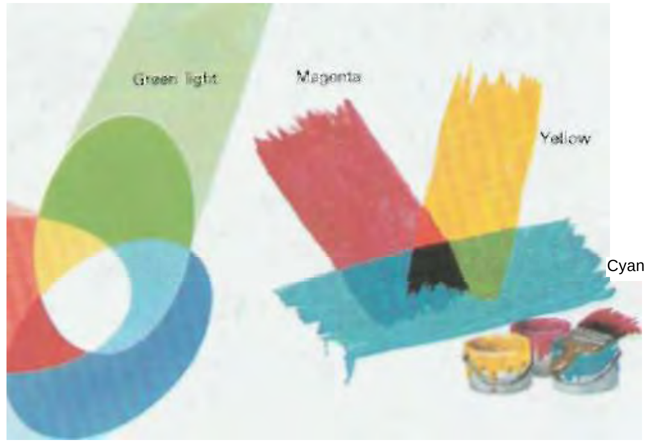
T *Light rays are reflected at surfaces in such a way that the angle of reflection always equals the angle of incidence. A plane mirror produces a virtual image by reflecting light so that it appears to the eye to come from behind the mirror. A convex mirror produces a virtual image, but diminished in size. The size of image produced by a concave mirror depends on the relationship between the position of the object and the focal length of the mirror- the point to which it converges light parallel to the axis. Here an inverted. reduced image is formed by a concave mirror; this image is also real and could not be seen by the eye.*



Newton claimed the rainbow contained seven colors, not because they were easily distinguishable but by analogy to the notes of the musical scale

What is white?
When Newton directed a beam of sunlight through a prism he found it split into colors varying from red to purple, as in a rainbow. In analogy with the seven notes in music (A to G), he defined seven colors - red, orange, yellow, green, blue, indigo, violet— though few people find it easy to recognize seven bands of color. The colors are light of differing wavelengths, varying from 700nm for the limit of the red end of the visible spectrum, to 400nm at the violet end. The color of a non-luminous object depends on the wavelengths of light that it reflects rather than absorbs. A white object is a perfect reflector, one that reflects all light; a black object is a perfect absorber.

Sources of light, such as the Sun and a tungsten filament light, emit a broad spectrum of wavelengths, which we perceive as "white". This contrasts with a sodium lamp, for example, which emits most strongly at two closely-spaced wavelengths in the yellow region. The continuous spectra from the Sun and a tungsten light vary in intensity in a manner characteristic of a perfect emitter, or "black body" (a perfect emitter is also a perfect absorber). The intensity rises to a maximum at a wavelength that depends on the temperature of the emitter. The Sun's radiation peaks around 500nm, corresponding to a temperature ofdOOOK. A tungsten filament light, on the other hand, runs at a temperature of about 2000K, and its spectrum peaks at 1500nm, well into the infrared part of the electromagnetic spectrum (^ pages 60-61). This is "heat" radiation; the visible light from the lamp comes from the higher wavelength, but lower intensity, side of the lamp's emission spectrum.



A *Mixing colored lights is an additive process. Red, green and blue together stimulate all three types of color responsive cell in the eye, and the result is seen as white. These primary colors of light can be mixed in pairs to give secondary colors. With a paint, the process of producing a color is subtractive -the pigment absorbs light at certain wavelengths /colors); the observed color results from removing the wavelengths from white light. Thus if 3 pigment absorbs red, it will reflect the secondary color made from combining the other two primaries (cyan).*

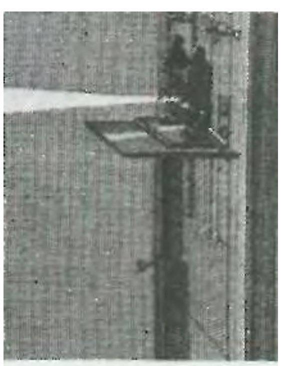
The color of light

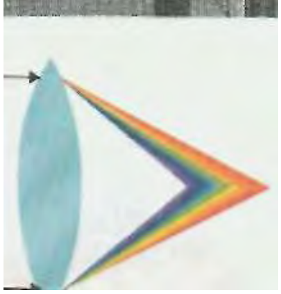
Refraction reveals another property of light - its color. By the 17th century, the ability of a glass prism to produce a broad spectrum of colors from a beam of "white" sunlight was well kown. However, it was the British mathematician and physicist, Isaac Newton (1642-1727) who made the first serious study of the nature of color. He proved for the first time that color is a properly of light itself, and has nothing to do with the nature of the prism or any other material. The prism "disperses" the light, refracting it according to its color. It refracts the red light the least and the violet light the most. Thus, the number quoted for the refractive index of a material depends on the color of light being used. For example, the refractive index for crown glass varies from 1 -524 for red light to 1-533 for blue light.

Objects appear colored because they absorb certain wavelengths and only reflect those which go to make up the color that is seen. A colored filter absorbs all light except those wavelengths that it allows to pass through. Thus a green object viewed through a red filler appears black - all wavelengths except green are absorbed by the object, and this green is itself absorbed by ihe filter.

Colors appear in a different manner when white light is reflected from thin layers of material, such as patches of oil or the outer "skin" of a soap bubble. The British physicist Robert Hooke (1635-1703) studied this effect in detail and discovered that the color observed depends on the thickness of the layer. However, it was only in 1801 that the basis for a proper explanation of this effect emerged. This was the concept of the "interference" of light (^ page 40).

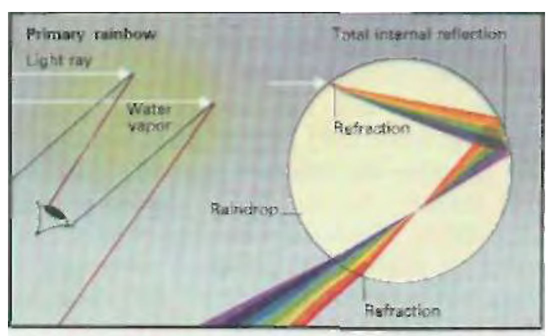
A *The British scientist Isaac Newton (1642-1727) made his fundamental discoveries about light and the nature ofcolor-and laid the foundations of his work on gravity and motion - while at home in Woolsthorpe in 1665-6 during the Great Plague, when the university at Cambridge was closed down. He published his first scientific paper on his work with prisms in 1672, and met with great controversy, particularly from Robert Hooke (1635-1703). Only when Hooke had died did Newton publish his work* Opticks *in 1704. Typicalofthe experiments Newton describes is one that splits white light into colors with a prism, recombines them with a lens, and then splits them again to form a spectrum on a screen. In this way he showed that colors are contained within white light.*

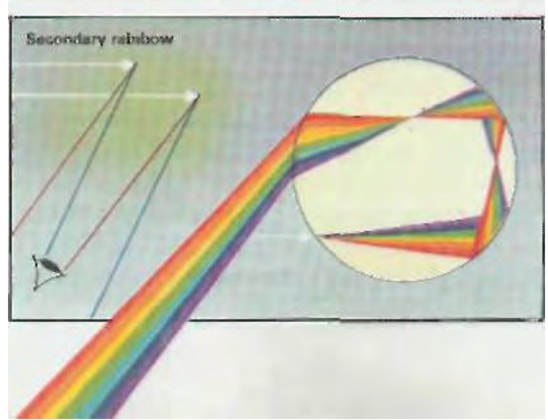




• Because the refractive index of a material varies with wavelength (color), a simple lens does not have a unique focus. Thus the lens forms a series of colored images of slightly different size, and the observed image appears to have a colored fringe. This effect - chromatic aberration - often occurs in inexpensive telescopes and binoculars.

A • *A spectacular demonstration of white light as a combination of colors is seen in a rainbow. Rainbows occur when sunlight from behind the observer falls on water droplets in front of him or her. Often a weaker "secondary" rainbow is seen outside the brighter "primary". In forming the primary, light from the Sun is first refracted as it enters a raindrop, then reflected from the back of the drop, and finally refracted again as it emerges, spread into the whole spectrum of colors. In forming the secondary rainbow, the light is reflected twice within the raindrop before it emerges. The additional reflection has the effect of reversing the order of the colors, so that although red appears on the outside edge of the primary bow, it is at the inner edge of the secondary rainbow.*





The wave theory of hght

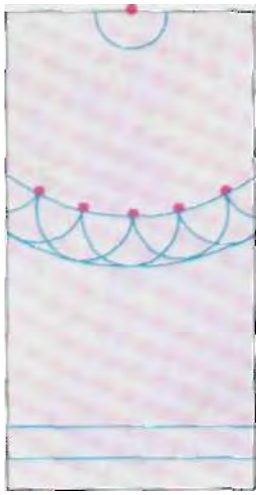
Thomas Young consolidated the idea that light is a wave motion, which he believed was "excited" in a "luminiferous ether (that) pervades the universe". Such ideas had been discussed by Hooke and others over a century before (| page 44) but Young was the first to recognize an important property not only of light but of wave motion in general. This is the "principle of superposition", which stales that when two (or more) waves cross, the size of the resulting wave at each point is given by simply adding together the sizes of the individual waves at that point.

This principle is revealed in an experiment first performed by Young, in which light falls on a card with two very narrow slits that are not far apart. The two narrow beams of light emerging through the holes illuminate a screen beyond. As the beams of light originate from the same initial beam, their wave motions should be in phase (undulating in unison). According to the superposition principle, at points where the light takes paths of different lengths to reach the screen, the separate undulations can be out of phase, as the trough in one wave arrives at the same time as the peak in the other. The two waves therefore cancel each other out to give darkness at the screen. This is confirmed by the pattern of alternating bright and dark stripes that appears on the screen. The bright stripes correspond to where the difference in the paths equals an exact number of wavelengths, so that the separate beams reinforce each other. The intermediate dark regions are where the two waves cancel each other.

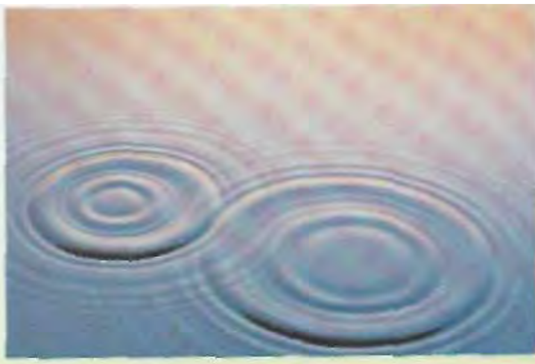
Such bright and dark bands occur when the two-slit experiment is performed with monochromatic light (of one spectral color). With white light, the pattern produced is a complex series of bands of different colors. This is because the reinforcing and canceling occurs at different points on the screen for different parts of the spectrum, and it shows that light of different colors has different wavelengths. The difference in the two paths necessary to reinforce (or cancel) the light must be slightly different for each color. Red light has the longest wavelength (requiring the largest path difference) while violet light has the shortest wavelength (smallest path difference).

The wave theory of light is the key to understanding color, and the concept of interference provides an explanation for the colors of thin layers. Light is reflected from both the top and the-bottom ^ of the layer, and the two reflected beams, of light can interfere exactly as when they emerge from two slits. The wave theory of lighT also explains the phenomenon of "diffraction", first noted in the 17th century by the Italian Jesuit scientist, Francesco--GrimaTdi (1618-1663). Grimaldi observed that the shadows cast by narrow-beams of sunlight in darkened rooms do not have precisely sharp edges but have colored fringes and these fringes spread beyond the expected edge of the shadow. ^—

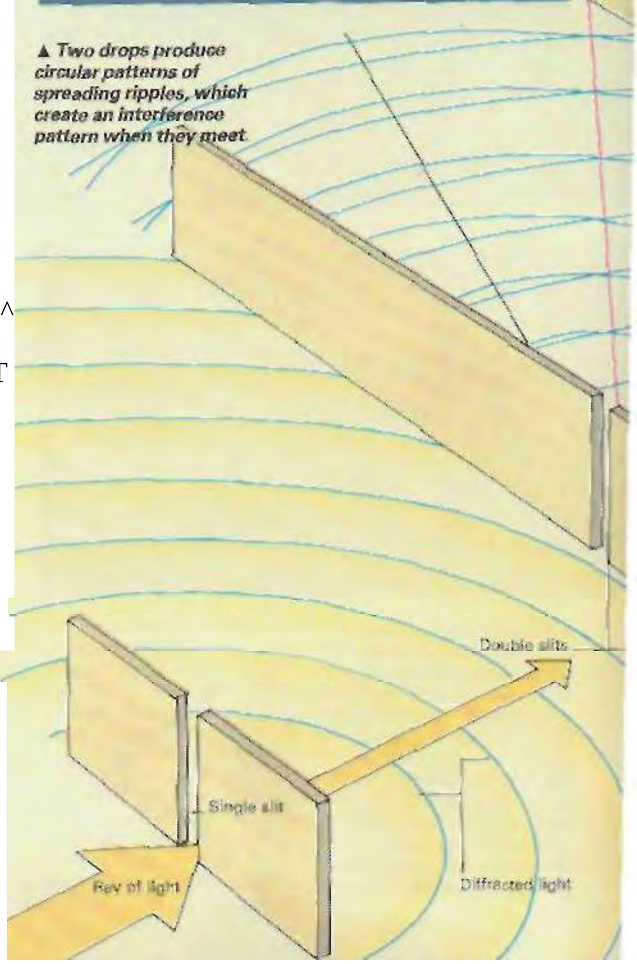
This effect is seen more clearly when monochromatic light passes through a single narrow slit to fall on a screen beyond. TheJight forms a pattern centered on a bright line, with a series of bright "fringes" gradually fading away to either side. Moreover, the central bright line is wider than the slit itself, the width of the line being inversely proportional to the width of the slit. The light apparently spreads as it emerges from the slit, as if the slit itself behaves like a row of little sources of light, each emitting a circular "ripple" of light. The fringes are caused by interference between the light waves from these "sources".

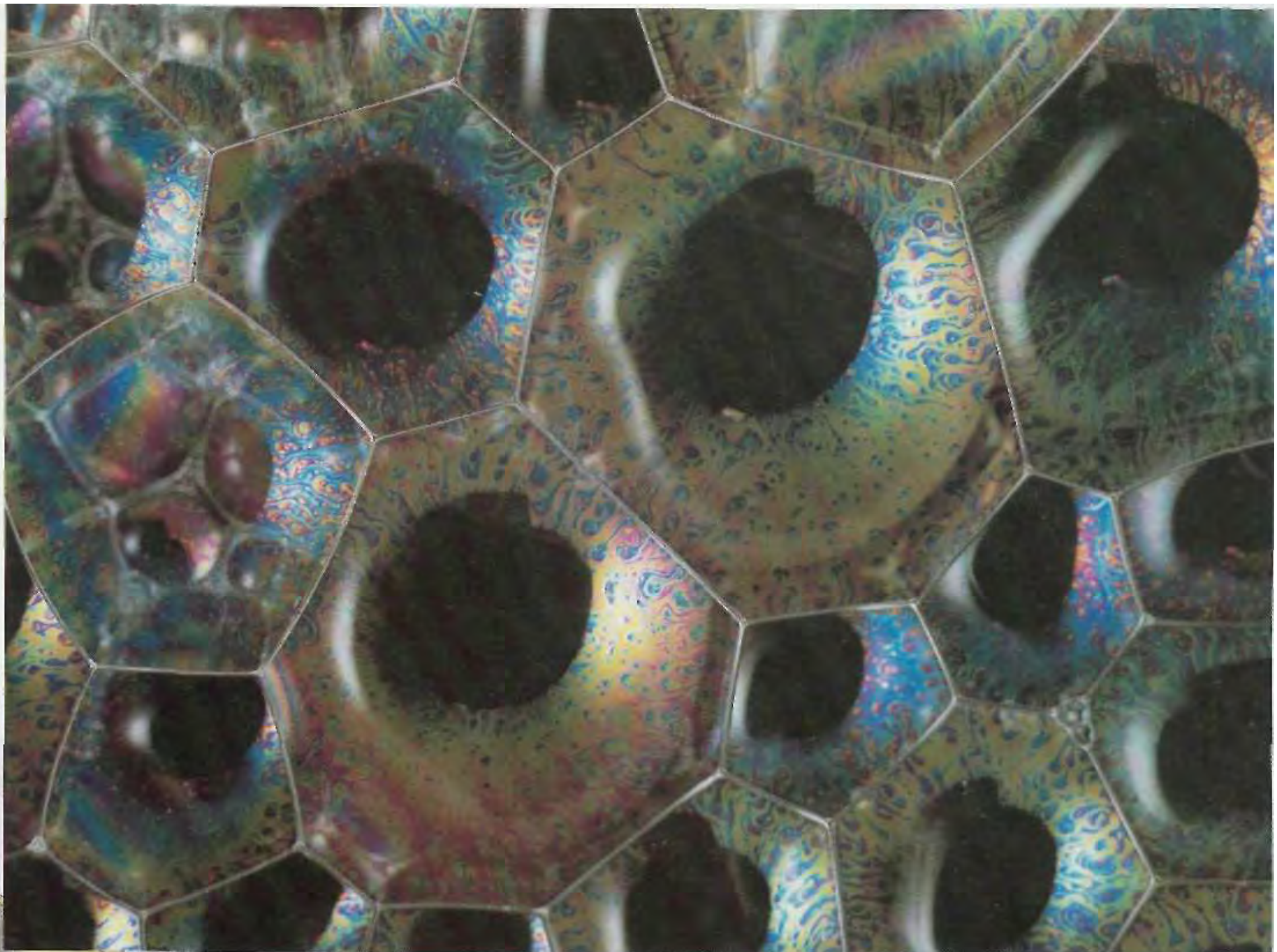


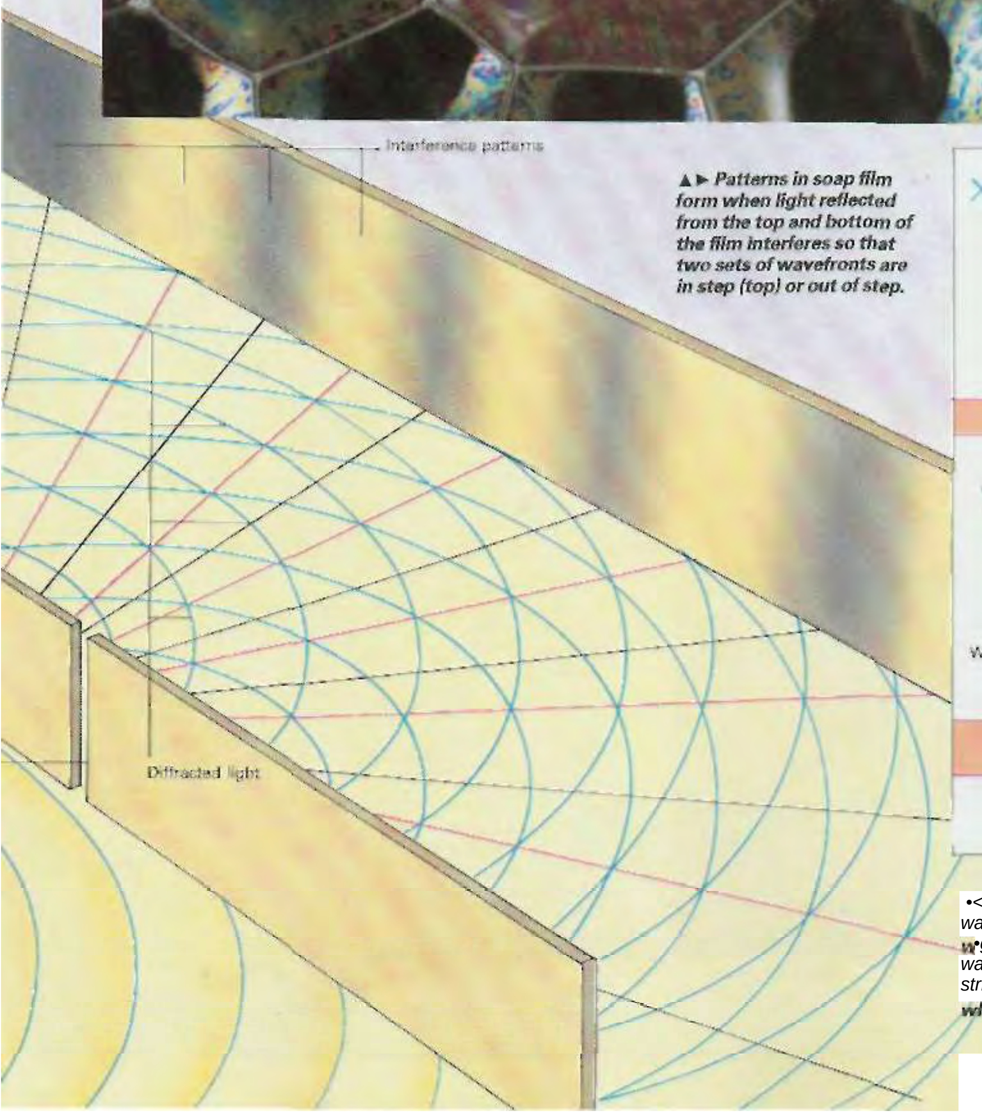
< The Dutch physicist Christian Huygens laid down the first foundations of a wave theory of light in 1678. He imagined that a point of light emits a spherical "wavefront", and that each point on this wavefront can be regarded as a new source of waves, and so on. The envelope of all the new "wavelets" gives the shape of the new wavefront, showing how the light spreads from the source. At large distances from the source, the wavefronts are in effect parallel. Huygens' principle successfully explained optical phenomena such as reflection and refraction, as well as interference.



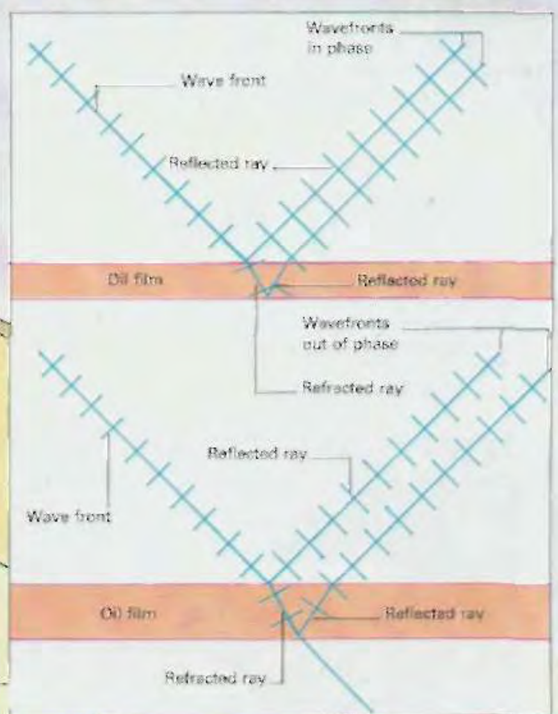
▲ Two drops produce circular patterns of spreading ripples, which create an interference pattern when they meet.







▲► ***Patterns in soap film form when light reflected from the top and bottom of the film interferes so that two sets of wavefronts are in step (top) or out of step.***



◄ Light shining through a pin hole produces spherical wavefronts which create two new secondary sources of wavefronts at a screen pierced by two holes. These new wavefronts interfere to produce a pattern of bright and dark stripes - bright where the wavefronts exactly match, dark where they are out of step and cancel each other out.

The Special Theory of Relativity

Einstein's two simple statements

If an experiment were conducted in a stationary laboratory on the Earth's surface, and then in an identical laboratory in a train moving with constant velocity along a level track, Newton's laws of motion would apply equally well in both instances. The velocity of the laboratory does not affect the results of the experiment. Indeed if the two laboratories had no windows it would be impossible to conduct any mechanical experiment to distinguish between them. This is known as the principle of relativity. All motion is relative - even the laboratory which is at rest relative to the Earth, is moving rapidly relative to the Sun. Yet Newton believed that there exists some point in space which is at "absolute" rest.

In the 19th century, the British physicist James Clerk Maxwell (1831-1879) demonstrated how electricity and magnetismare related, and showed that light is an electromagnetic wave 0 page 61): Scientists assumed that waves could not travel through empty space so the existence throughput space of a transparent medium, called the ether, was suggested. This could conceivably be in the state of absolute rgst suggested by Newton. Researchers conducted a number of experiments to find the velocity of the Earththrough the ether. In 1887, two American scientists Albert Michelson 11852-1931) and Edward Morley (1838-1923) used an optical device in such an attempt, but, despite the accuracy of their measurements, they were unable to obtain a result. Among the explanations for this was a suggestion that any object moving through the ether suffers a "contraction " in the direction of motion.

In 1905, Einstein proposed his special theory of relativity, based upon two simple statements that, when taken together, lead to some extraordinary conclusions. First, he stated that no physical experiment, including optical ones, can distinguish between the two identical laboratories mentioned above; consequently there is no point in trying to detect the absolute motion of the Earth as Michelson and Morley had attempted. Second, he stated that all observers must obtain the same value when they measure the speed of light through space, even though the source of light may be approaching them or receding from them. Although supported by Michelson's and MorleYs experiment, this contradicts our usual experience - a stone thrown forwards from a moving automobile moves faster relative to the ground than one thrown from a stationary vehicle.

Some of the consequences of these statements can be seen by considering two spacecraft, A and B, with B moving past A at half the speed of light. Some remarkable conclusions ensue.

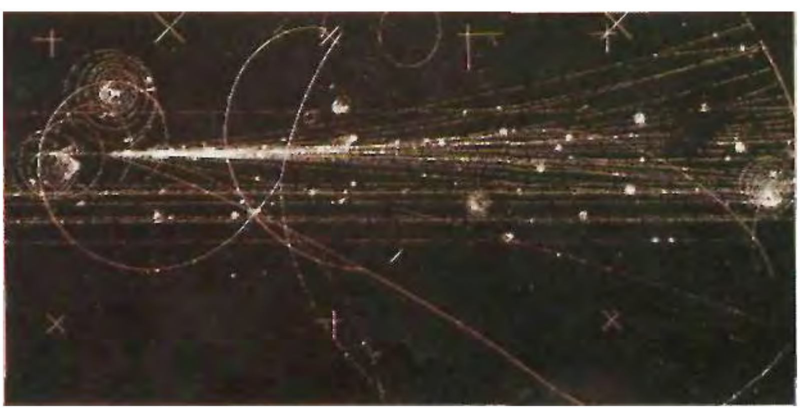
A sees B's spacecraft Shortened in the direction of travel. In fact a meter-rule on B's craft appears to be only 87cm long. If A were able to measure the mass of B as he passes in his craft he would discover that it had increased. His 75kg colleague would appear to have a mass of 86kg.

Einstein showed that the relative mass increase of an object was a measure of the energy imparted to it. He argued that if the additional mass is a measure of energy then surely all of the mass should have an equivalent energy value. This equivalence between mass and energy is expressed in the relationship, energy equals mass times the velocity of light squared ($E=mc^2$). A would also observe that B's clock runs slower than his. None of these effects will be noticed by B, however, in his view nothing has altered on his craft, but A (receding at high speed) seems subject to all these changes.

• The special theory of relativity rests on two basic postulates. The first is that the laws of physics are the same in all inertial (non-accelerating) frames of reference. Thus passengers on a high-speed train can pour coffee and walk about with no awareness of their rapid motion unless they look out of the window.

• Einstein's special theory of relativity set out to agree with the experimental finding thatthe velocity of light does not depend on the motion of the observer. In so doing, the theory made several remarkable predictions. Two of these predictions concern changes in the observed mass and length of an object as its speed increases towards the velocity of light. Lengths parallel to the direction of motion (but not those at right angles to it) decrease, or contract, while masses increase. These changes in length and mass make it impossible for anything to travel faster than the velocity of light.

Vtt,

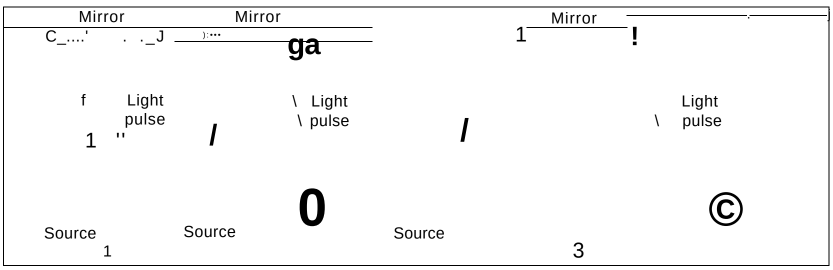


•4 An object's mass is a form of energy. In this picture of tracks in a bubble chamber, the collision of a single high-energy proton with a stationary one creates a spray of new particles, as kinetic energy converts to mass energy.

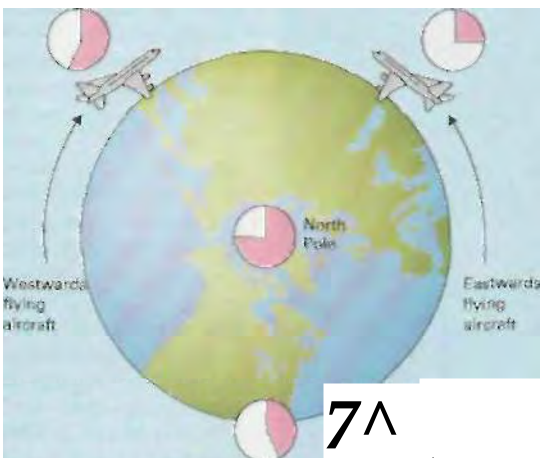
0.2 0.3 0.5 0.6 0.7 0.9

Proportion of speed of light

Speed of lighi.



A *The second postulate of the special theory is that the speed of light is independent of the velocity of the source. One of the consequences of this is the phenomenon of time dilation. Suppose a moving clock sends out a light pulse while coincident with a stationary clock, and the light is reflected back from a mirror 11). From the viewpoint of the stationary clock the light will appear to have traveled farther to the moving clock, so that a stationary observer will say that the time interval for the moving clock has increased (2). The faster the clock is moving, the greater the dilation will seem (3).*



k *Two clocks were flown around the world, calibrated against a clock in Washington. After returning, the eastward-flying clock had lost time, and the westward gained, relative to the clock in Washington. The explanation requires an imaginary clock at the North Pole, stationary relative to the rotation of the Earth. The clock in Washington is moving relative to this clock, and so loses, whereas the eastward-flying clock, moving faster than the Earth's rotation, loses more time. The westward-flying clock, moving against the Earth's rotation, seems from the Pole to be moving slower than the clock in Washington, so, though losing time relative to the Pole, it gains relative to Washington.*

Albert Einstein

Albert Einstein was born in Ulm in Germany in 1879. He attended a Catholic grammar school in Munich, despite being a Jew, but showed no great promise and later took himself to Switzerland to continue his studies. There he missed most of his lectures, and relied on the notes taken by a friend. When he graduated in 1901 Einstein became a Swiss citizen and obtained his first job as a junior official in the patent office in Berne.

In 1905 Einstein had a remarkable year, publishing no fewer than five important papers. In one he explained the photoelectric effect, thus laying the foundations for quantum theory (t page 88); in another he provided the first satisfactory explanation for Brownian motion I'f page 27); but perhaps most famous is the paper describing the special theory of relativity. This dealt only with unaccelerated motion, and in 1915 Einstein published his general theory of relativity which covered the more general case of accelerated motion. This theory contained a new description of gravity which replaced and extended Newton's theory (} pages 108-109).

Einstein was appointed to professorships in Zurich and later in Berlin. In 1922 he was awarded the Nobel Prize for his work on quantum physics. He emigrated to the United States when the Nazi Party came to power in Germany in 1933 and it was from the Institute of Advanced Studies at Princeton, New Jersey, that he wrote to President Roosevelt urging the construction of the atomic bomb. However, after World War II, Einstein argued tirelessly for a world agreement to end the threat of nuclear war. Unfortunately his influence was reduced by his habit of signing almost every petition that was put in front of him. He died on 18 April, 1955, in Princeton.

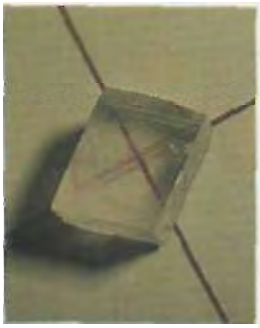
Flying clocks around the Earth

One of the most convincing tests of special relativity was performed in 1971, when two United States physicists, Joseph Carl Hafele and Richard Keating, flew around the world with four atomic clocks. They made two journeys - on passenger jet aircraft- one eastward and one westward, and both taking about three days. By comparing their clocks at the start and finish of the journeys with reference clocks in Washington, Hafele and Keating observed how well their clocks had kept time.

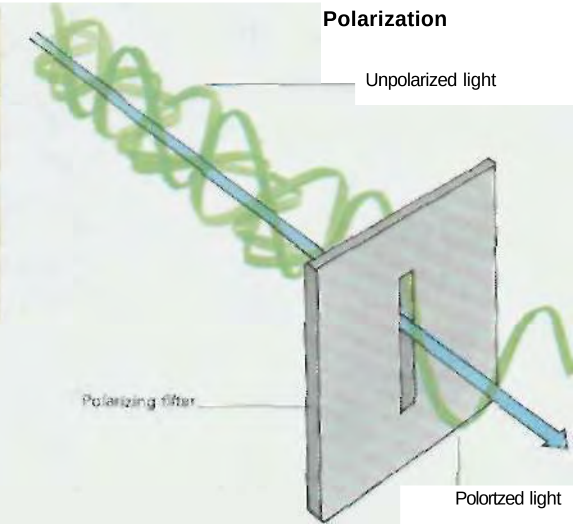
They found that the eastward-moving clocks had lost 59 billionths of a second compared with the reference clocks in Washington, whereas the westward-moving clocks had gained 273 billionths of a second. The clocks had been influenced by two effects. First, both sets had gained because they were flying at high altitudes where the gravitational force of the Earth is weaker and time runs more quickly - this is an effect predicted by Einstein's general theory of relativity. However, the eastward-moving clocks had also lost time because their velocity was greater relative to the clocks on Earth. Indeed, the eastbound clocks lost more time than they gained. The westward-moving clocks, however, gained still more relative to those on Earth, because their velocity was less than that of the clocks in Washington.

See also
Sound 21-1
Zlectromagnetism 57-64
Tho Quantum World 87-96
Fundamental Forces 105-10

Polarization

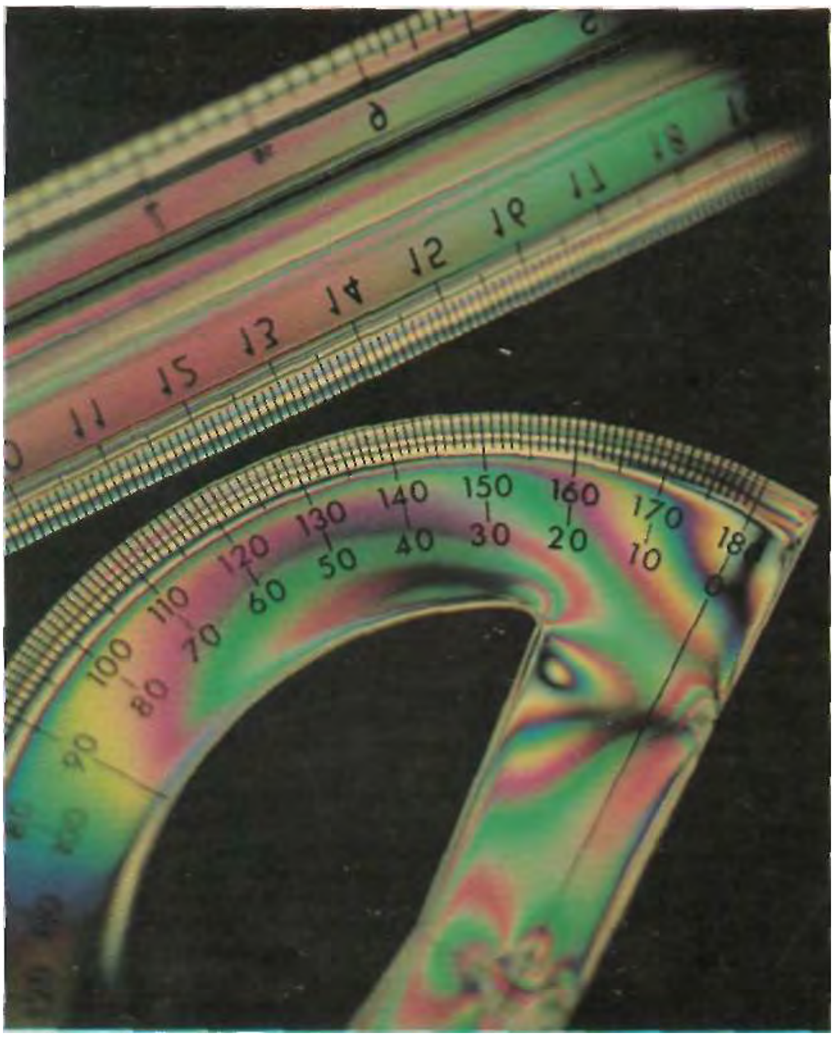


• Calcite forms two images in certain directions as the crystal splits light into two rays. One ray refracts normally; the other, which here forms the displaced upper image, does not. In the line at right angles, the images are superimposed.



A Ordinary - unpolarized - light consists of vibrations in all directions at right angles (transverse) to the direction of travel, but these can be resolved into two directions at right angles. A polarizing filter transmits light only in one of these directions, so that it becomes "plane" polarized. This effect is used in polarizing sunglasses, which transmit only the light plane polarized in one direction.

J Mechanical stress makes acrylic doubly refracting, so it splits light into two rays with different polarizations and velocities. This causes colored interference patterns in polarized light.



Particles, waves and polarization

In 1669, the Danish scientist Erasmus Bartholinus (1625-1698) acquired some crystals of calcite, or "Iceland spar", which had been brought to Copenhagen from Iceland. Bartholinus discovered that he could see two images through the crystals, and he concluded that the crystal split an incident ray of light in two - a process of "double refraction ". One of the rays obeyed Snell's law of refraction (4 page 36), the other did not. Another Dutch physicist, Christian Huygens (1629-1695) took these investigations further, using a second crystal to study the two beams of light emerging from the first crystal. He found that either one of the two images would become more feeble and disappear as he rotated the second crystal.

Huygens also developed a wave theory of light, first published in 1678, in which he proposed that light propagated in the form of spherical waves. Moreover, he suggested that each point on an advancing spherical "wavefront" is the source of a new spherical wave (4 page 40). According to Huygens, this was how the waves propagated, through a "very subtle and elastic medium " that he presumed to fill all space. Huygens' theory proved successful in explaining reflection and refraction. Indeed it was an inspiration to Young more than a century later, and even now provides the simplest way to understand diffraction patterns. However, Huygens could not explain properly all the effects observed with the calcite crystals.

At the time of Huygens, however, the concept of light as a wave motion lost out to the alternative idea of light as beams of particles, which dominated thinking about optical effects for the next 100 years or so. Newton is often regarded as stressing the corpuscular, or particle, nature of light, but he seems to have preferred a bizarre mixture of corpuscular and wave theories, in which light is emitted as particles which then set up vibrations in the "ether" (equivalent to Huygen's "elastic medium "I pervading all space. But even Newton was unable to explain the behavior of calcite. He postulated that rays of light have "sides", and in a sense foreshadowed what is now referred to as the "polarization " of light.

Young reintroduced the wave theories in 1801, but his ideas were at first criticized; it fell to the French physicist Augustin Fresnel (1788-1827) to show that all the optical phenomena known at the time could be understood in terms of light waves, including the double refraction of calcite crystals.

In 1808, Etienne Malus (1775-1812), an officer in the French army, was experimenting with some Iceland spar when he noticed that if he put the crystal in a beam of light reflected from a window, then the two images would vary in intensity just as Huygens had first observed with two crystals. Malus had discovered that light acquires "sides" when it is reflected. Nine years later Young pointed out to Fresnel's colleague, Dominique Arago (1786-1853), that the "polarization" of light upon reflection could be explained if it was assumed that light vibrates transversely to its direction of motion. However, the physical meaning of these transverse vibrations was not explained until Maxwell's theory of electromagnetic waves ($ page 57).

Magnetism

Forces and fields...The nature of magnetism...Magnetic domains...PERSPECTIVE...Early studies in magnetism... The Earth as a magnet..Compasses...Using magnetism for storage...Magnets and medicine

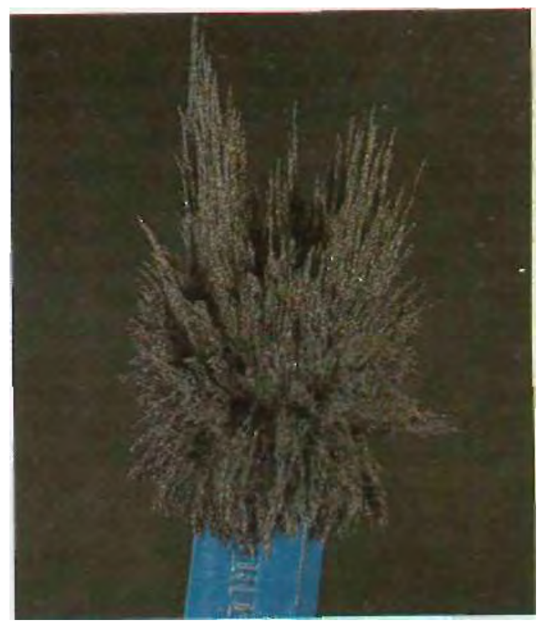
k Magnets used in scrapyards are electromagnets, in which an electric current induces a magnetic field in iron. The induced magnetism does not last long once the current is switched off, and the magnet can release its load.

A magnet is a piece of iron or other material that can attract or repel other similar materials placed nearby. If a bar magnet is suspended away from all other magnetic materials so that it can rotate in a horizontal plane, it will always tend to align itself the same way, which by convention is called the north-south direction. The end that points north is called the north (or north-seeking) pole, and the other end is the south (or south-seeking) pole. All magnets have two poles; a single magnetic pole has never been found despite speculations that such objects may exist (^ page 96).

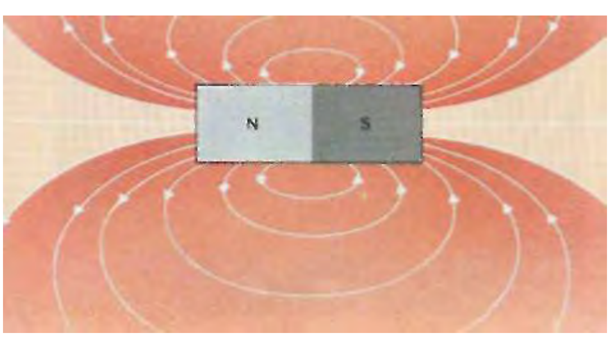
The magnetic field

The region round a magnet over which it exerts a force is called the "magnetic field". This field can be represented as a series of lines in three dimensions that form a pattern finking the north and south poles of the magnet. Conventionally the lines represent the way in which a (hypothetical) free north pole would move, its direction shown by an arrow. The field is strongest at the poles, and decreases with increasing distance from them. Like the electrostatic field (^ page 49), it decreases in inverse proportion to the square of the distance. In the pattern of field lines, the fines are closest where the field is strongest.

When a piece of unmagnetized magnetic material, such as iron, is brought near to a magnet it becomes temporarily magnetized - that is, north and south poles form - and it is attracted towards the magnet. This effect is called "induced magnetism". The strength of the induced magnetism and how long it lasts depend on the properties of the material. Iron is a "soft" magnetic material: it can be magnetized and demagnetized easily, and forms a strong induced magnet, but its induced magnetism does not last for long. Steel is a "hard" magnetic material: it is more difficult to magnetize strongly but retains its magnetism for longer. Steel is therefore used for making permanent magnets, whereas iron is used for electromagnets, where the magnetism is switched on and off by an electric current (^ page 57).

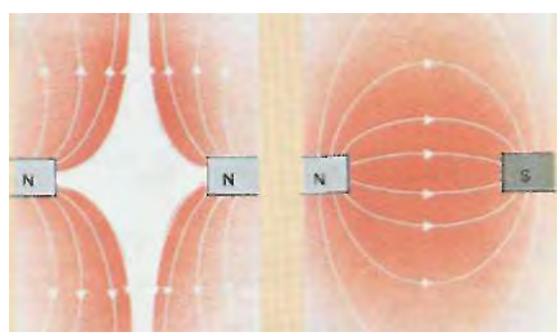


A Iron filings reveal the field on the end of a magnet.

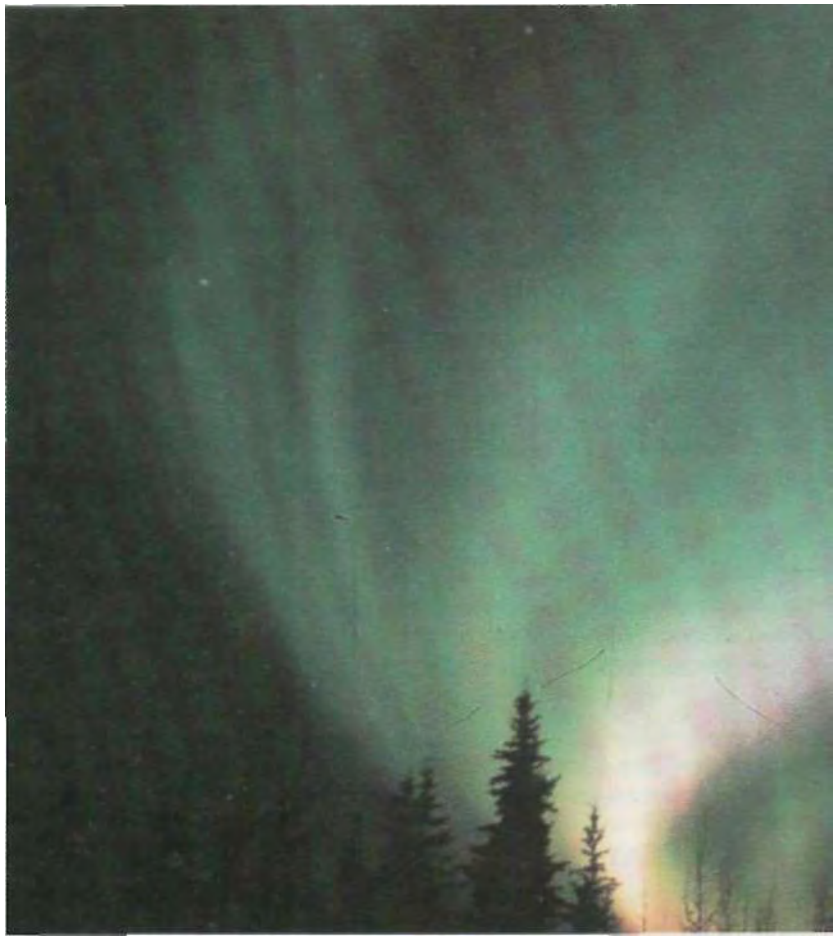


•4 A magnetic field can be represented by lines linking north and south poles in a three-dimensional pattern. The arrows give the direction of the force on an imaginary north pole.

• Similar magnetic poles repel each other (right), giving a "neutral" point between them where there is no net force. Dissimilar poles (far right), on the other hand, are attracted to each other, with the field lines linking them together.



The reason why the Earth has a magnetic field has never been fully demonstrated



The Earth as a magnet

The observation that magnets align themselves along a north-south axis implies that the Earth itself must possess a magnetic field. Indeed, the Earth behaves as if it has a bar magnet at its center. Magnets and compass needles align themselves with the Earth's lines of magnetic force. These lines are not exactly along geographic meridians, so the north and south magnetic poles do not coincide with the geographic ones. Their positions are slowly changing all the time, so the declination (the angle between magnetic and geographic meridians) is variable.

If a compass is free to rotate in a vertical plane as well as a horizontal one, it dips towards the surface of the Earth at varying angles depending on latitude. At the north and south magnetic poles a dip needle points vertically while at the magnetic equator it lies horizontally.

Physicists have deduced from a study of earthquake waves that the Earth has a liquid core; they are therefore confident that there is no real bar magnet under the Earth's surface. One possible explanation for the Earth's magnetic behavior is that the liquid core has circulating electric currents inside it, which make it behave like a gigantic dynamo (+ page 59).

The Earth's magnetic field extends some 80,000km into space and this causes many strange effects such as electric winds and aurorae, when the field traps charged particles from the Sun and outer space.

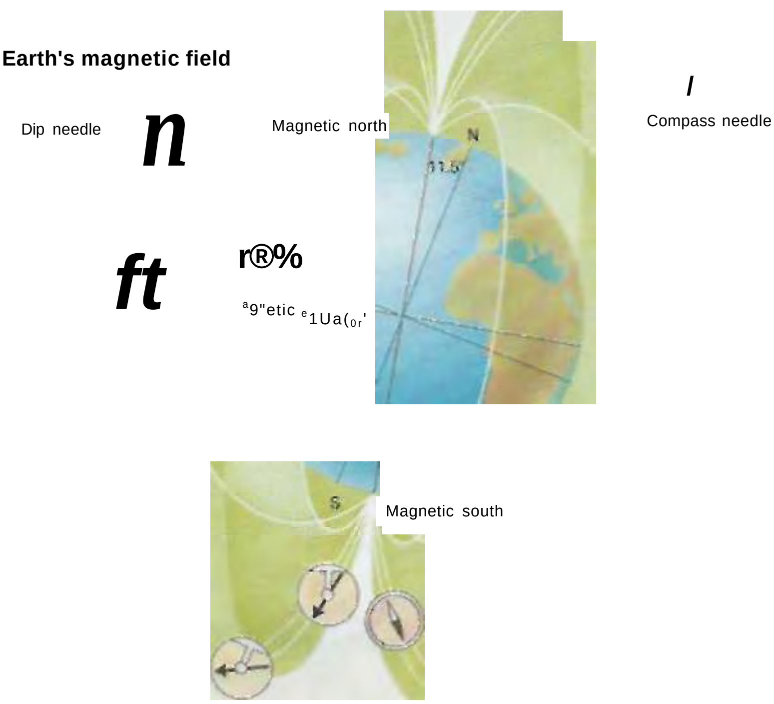
The Earth's magnetic field periodically reverses itself (that is north and south poles change places). This has been discovered in studies of magnetized rocks of various ages, some of which have a polarity opposite to the present Earth's field. Because the magnetization was produced when the rocks were formed it is possible to draw up a timechart of polarity changes. This, together with information relating to polar position, intensity of field, and angle of inclination, can then be used for dating rocks and determining their origin.

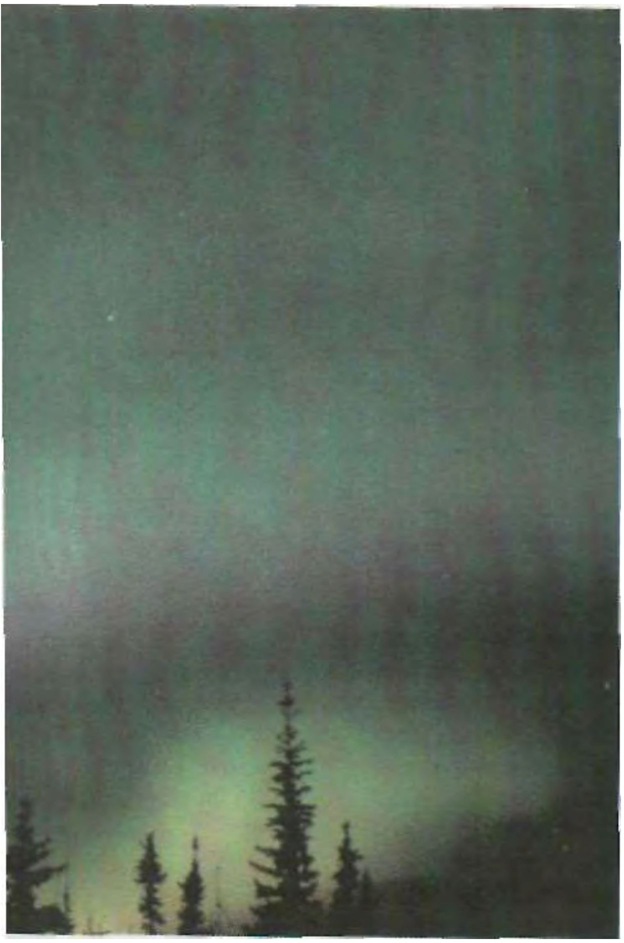
T *The Earth has a magnetic held similar to that of an imaginary bar magnet located at its center, with the magnet's south pole in the Northern Hemisphere. Magnetized needles free to rotate about a pivot align themselves with the Earth's magnetic field, which is horizontal at the magnetic equator and vertical at the magnetic poles. The magnetic and geographic poles are usually at slightly different locations.*

T *The value of compasses has been known since at least the 12th century. The main component is a magnetized needle pivoted about its center and free to rotate in the horizontal plane. In the absence of other magnetic fields, it will align with the horizontal component of the Earth's field. Instruments known as dip circles, in which a needle rotates in a vertical plane, indicate the vertical component of the field.*

A *The Earth's magnetic environment becomes visible in spectacular fashion in the displays of Aurora Borealis (in the Northern Hemisphere) and Aurora Australis (Southern Hemisphere). Aurorae occur at altitudes of 100-300km when electrons spiral down the Earth's magnetic field toward the poles. The electrons excite molecules, which emit light at different colors as they return to the ground state It page 88).*

Earth's magnetic field





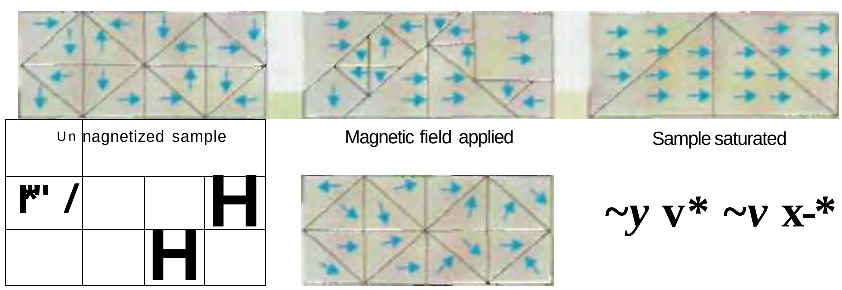
Magnetism is intimately related to electrical effects. In 1820 the Danish physicist Hans Christian Oersted (1777-1851) discovered that an electric current flowing through a wire sets up a magnetic field around the wire «> page 57). If the wire is twisted to form a coil, then the shape of the magnetic field outside the coil is similar to that of a bar magnet. The strength of the field can be increased by raising the size of the current or the number of turns of wire in the coil. The field strength can also bc increased by placing a core of soft magnetic material inside the coil. Such a coil is called an electromagnet.

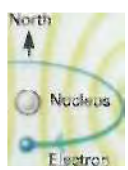
How does a material become magnetized?

In his electromagnetic theory (| page 61), the British mathematician and physicist James Clerk Maxwell (1831-1879) proved that moving electric charges - that is, electric currents - are the source of magnetic fields. The electrons and protons that constitute atoms are moving electric charges and are therefore themselves sources of magnetic fields. Different materials exhibit different magnetic properties and respond in different ways to external fields. Such differences can be traced to the electronic configuration and arrangements of the atoms. In magnetic materials the atoms form groups called domains, each of which contains many molecules. In a domain, the "atomic magnets" line up so that the assembly has a net magnetization. In general, neighboring domains are magnetized in opposite directions. However, the application of an external magnetic field encourages the domains aligned to this field to grow in size at the expense of their neighbors. In the absence of the external field, the agitation of ihe aioms and (he forces at the poles disturb the alignment of the domains.

Only a few materials (iron, nickel, cobalt and some alloys and rare-earth elements) show any permanent magnetism. These are called ferromagnetic materials and become magnetized in the direction of even a weak external field. "Paramagnetic" materials become weakly magnetized in the same direction as a very strong external field. "Diamagnetic" materials become weakly magnetized in the opposite direction to a very strong external field. Indeed, all materials exhibit diamagnetic properties, but usually these are obscured by para-magnetic or ferromagnetic behavior.

Magnetic domains





•4 A Magnetic fields are created by moving electric charges as Maxwell showed in his theory of electromagnetism. The magnetism of materials is therefore related to the basic structure of atoms, with negative electrons orbiting positive nuclei. In a magnetic material, the atomic "magnets" line up in small regions called domains, which are usually oriented at random, giving no net field. When an external magnetic field is applied, domains which are already aligned with the external field grow at the expense of others (top), so that the material becomes magnetized. In strong fields (below), the domains rotate direction to become aligned.

Early studies in magnetism

The attractive and repulsive properties of naturally magnetic rock, lodestone or magnetite (Fe_3O_4), were probably known in prehistoric times, but it is believed that the Chinese discovered its directional properties. As early as 2500 BC a piece of magnetite was supposedly used as a primitive compass by a Chinese emperor to guide his troops through fog.

The Greeks were aware of the attractive properties of lodestone, and their literature contains fables referring to mountains that could draw nails out of ships or make it difficult for a shepherd (with iron tacks in his shoes) to move. By tradition the Greek philosopher Thales (624-546 BC) was the first to observe and study this effect.

It is not certain how or when Europeans discovered the directional properties of lodestone. By the 12th century there are references to the use by navigators of a compass consisting of a needle-shaped magnet floating on a reed or splinter of wood in a bowl of water. The more familiar form of compass, with a needle resting on a pivot over a circular scale, is due to the French scholar Petrus Peregrinus (c. 1240-unknown).

Peregrinus also determined north and south poles and found that like poles repel and unlike attract. He also broke magnets in half, and found that each half had two poles. However, he believed that the geographical poles were in the heavens and did not know that the Earth itself is magnetized.

The English physicist William Gilbert (1544-1603) showed that a compass needle dips towards the Earth, and suggested that the Earth itself is a great magnet. His researches into the properties of lodestone, and the differences between magnetism and static electricity (^ page 51) were extensive, but poorly followed up until the 19th century.

See also
Forces, Energy and Motion 11-20
Electricity 49-56
Electromagnetism 57-64
The Quantum World 87-96

Magnetism for storing information

An oscillating electrical signal can be used to create a varying magnetic field between the poles of a narrow-gap electromagnet. In this fact lies the principle of storing sound, images and computer data on magnetic media. The storage medium usually consists of finely-divided particles of iron or chromium oxide mounted on a tough plastic substrate in the shape of a tape or disk. As the medium passes the small gap (or tape head) the varying magnetic field induces varying magnetization in the metal oxide particles. The system works in reverse during playback.

Audio signals can be stored by driving a tape past a stationary head at a speed of about 5cm/s, although better quality recordings are produced at higher tape speeds. Video signals have higher frequency components and therefore require a much higher scanning speed. This is achieved by moving the tape slowly past a fast moving tape head, and results in the picture signal being stored in diagonal stripes across the tape. Computer data is in digital form with the signal either "on " (high magnetization) or "off". Such signals are pure and contain no "noise " due to the - demagnetization of the tape with time. Such techniques are now also being used in the digital recording of sound and video.

r In this experimental medical use of magnetism, a tumor cell in human bone marrow is surrounded by magnetic "microspheres" to allow doctors to draw the diseased cell from the body quickly, efficiently and without the disturbance to the body that irradiation or chemotherapy might cause.

•• Tiny particles of iron or chromium oxide store the information contained in sound waves as varying patterns of magnetization on the tapes used in audio cassettes. Magnetic effects also play a part in the transduction of sound and electricity in microphones, speakers and headphones.

Magnetism and the human body

In recent years doctors have found increasing applications of magnetism as an aid to treatment of disease; the fact that relatively strong magnetic fields can penetrate the body without damaging living tissue has proved a great advantage in this field. Nuclear magnetic resonance (» page 93) takes advantage of this fact to permit a cross-sectional or three-dimensional image to be created of the internal organs.

In the mid-1980s a new technique was developed in Britain to assist doctors in the treatment of cancers, particularly tumors in bone marrow. This technique takes advantage of so-called monoclonal antibodies, cells originating from the human immune system but processed by genetic engineering techniques so that they are "targetted" to attach themselves to specific cells in the human body. Monoclonal antibodies, designed to locate the tumor cells, are coated in tiny magnetic microspheres and inserted into the diseased bone marrow. As a result, the tumor cells effectively become responsive to a magnetic field, and can be drawn out of the bone marrow by application of a magnet. This technique, known as magnetic depletion, has proved highly effective in clinical trials, and may become an important form of treatment in the future.

Electricity

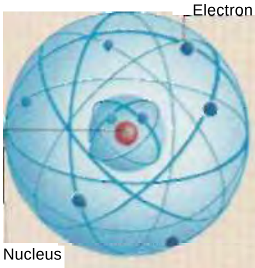
Electric charge, a fundamental of all matter...Attractive and repulsive forces...Field lines...Potential difference... Current and resistance... Capacitance...PERSPECTIVE... Static electric'rty...How a battery works... Semiconductors—Electrolysis

The world is built from atoms created from a small number of types of elementary particle. The complexity of the world results from the ability of these particles to affect the motion of other particles through fields of force that surround them. It is these interaction's that ultimately produce the familiar structures of the everyday world. Beyond the scale of tiie atomic nucleus, there are only two forces that cause these interactions: the electromagnetic force and the gravitational force O page 107). The gravitational force is very weak on the scale of particles: the electromagnetic force, on the other hand, determines the structure of atoms, of solids, liquids and gases, and of the complex molecules in living matter.

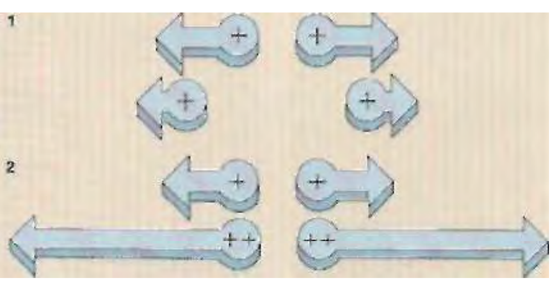
Most elementary particles are surrounded by an "electric field". Such particles are said to carry an electric "charge". The field is proportional to the strength of the charge, but falls off with the square of the distance from the charge. This "inverse square law" is named for the French engineer Charles Augustin Coulomb (1736-1806), who tested its validity in the 1780s.

The electric field due to a charged particle has a direction as well as a strength. The force between identical particles is repulsive, and it acts along the line joining the two particles. The field of each panicle exerts a force on the other to push them apart.

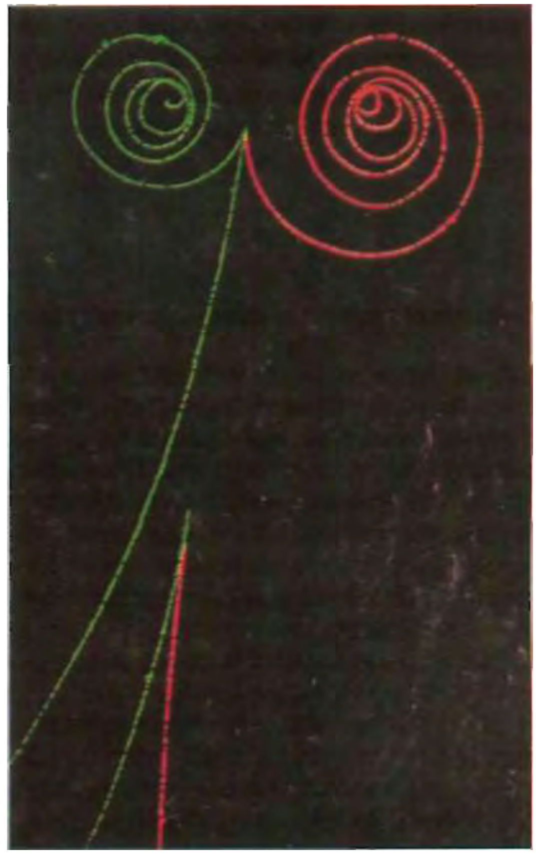
There are also, however, attractive forces that bind particles together. The electric force can attract as well as repel because there are two kinds of electric charge. These are labeled "positive" and "negative". Whereas two charges of the same kind repel, two opposite charges attract. In an atom a positively-charged nucleus is surrounded by negatively-charged electrons, which are attracted to the nucleus but repelled by each other O page 68). Overall the atom is electrically neutral. The fact that an atom contains electric charge is important in forming molecules from a number of atoms.



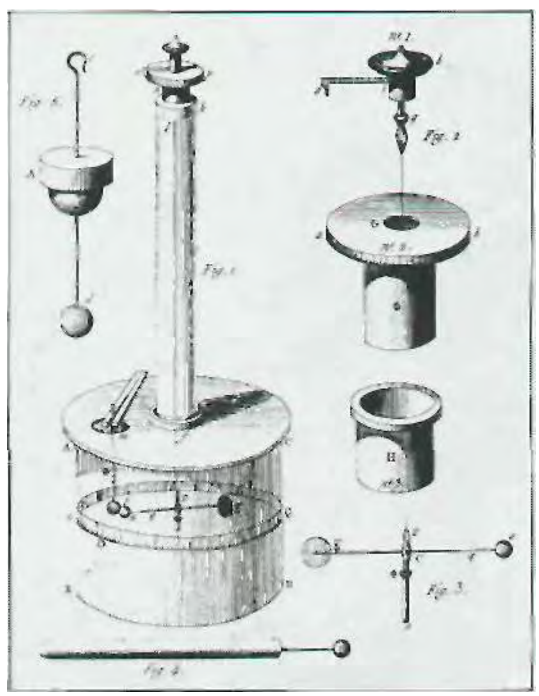
A The origin of electricity lies in the structure of matter. Matter is made up of atoms. Each is neutral overall, with negatively charged electrons orbiting a positive nucleus.



A • In the 1780s, Coulomb, using a torsion balance (right), established the basic law of force between two charged particles - the force decreases in proportion to the square of the distance between the charges (1), and it is directly proportional to the product of the charges 12). Double the distance between the charges and the force falls by one quarter; double the charges and it increases by a factor four.



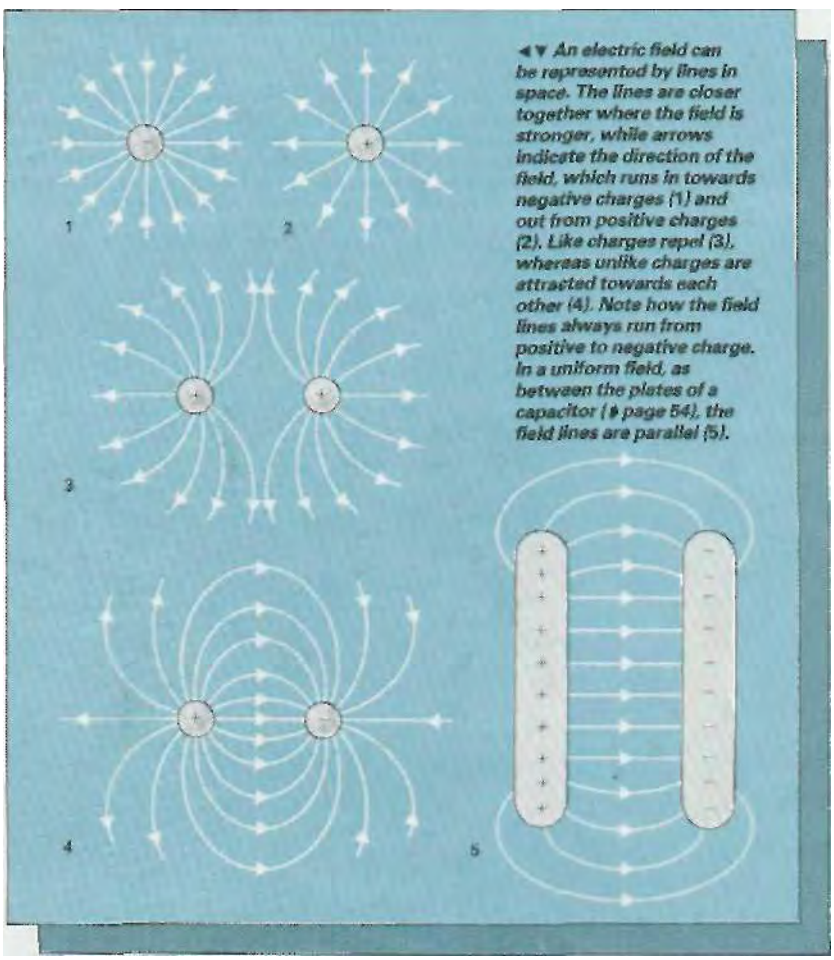
A Electric charge exists in two varieties - positive and negative - which behave in opposite ways. Here, particles with identical mass but opposite charge move under the influence of a magnetic field in a bubble chamber (t page 98). The particles are electrons (green tracks), each carrying one unit of negative charge, and positrons (red tracks), which are antielectrons and carry positive charge.



The word electricity derives from the Greek name for amber - elektron

The British scientist Michael Faraday (1791-1867) proposed a method of visualizing electric fields as lines in space. The lines point in the direction of the field, and are closer spaced where the field is stronger. Such imaginary field lines wind continuously through space, beginning only on a positive charge and ending only on a negative charge. The number of lines starting from, or ending on an electric charge, is proportional to the strength of the charge. These are properties unique to a field that obeys an inverse square law. The "electric potential difference" gives a measure of the amount of electric field between two points. For a constant field, the potential difference grows as the separation between the points increases, and it is proportional to the strength of the field. Electric potential difference is familiarly known as "voltage difference" or "voltage". Its unit is the volt (V), named for the the Italian physicist, Alessandro Volta (1745-1827), who invented the electric battery in 1800.

Voltage is a measure of work done (4 page 18), in this case the work done in moving a positive charge of one unit against the electric field between two points. This is like work done in walking from the bottom of a hill to the top, although in that case the work is done against the Earth's gravitational field. If the electric field between the points varies with position, or if the path is not a straight line, it is still possible to compute the voltage by dividing the path into tiny pieces and adding together the voltages across each piece. Such calculations show that the potential difference between two charges does not depend on the chosen path.



◄▼ *An electric field can be represented by lines in space. The lines are closer together where the field is stronger, while arrows indicate the direction of the field, which runs in towards negative charges (1) and out from positive charges (2). Like charges repel (3), whereas unlike charges are attracted towards each other (4). Note how the field lines always run from positive to negative charge. In a uniform field, as between the plates of a capacitor (▶ page 54), the field lines are parallel (5).*

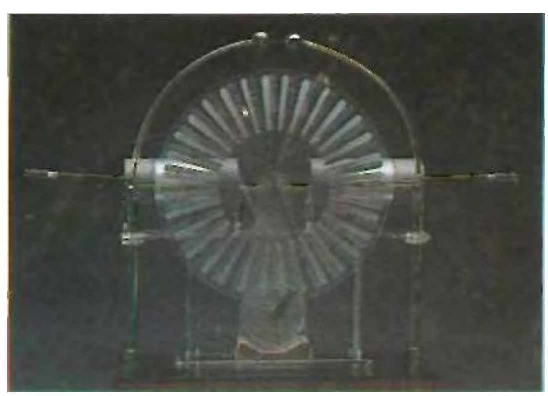
A *Experiments into static electricity were conducted around 1700 by the British physicist Stephen Gray 11666-1736). The "charity boy" (orphan) was suspended off the floor and charged electrostatically. Pieces of paper were attracted by his positive charge (depletion of electrons).*

Static electricity

Electric charge may be deposited on an object only by transferring charge to it from some other piece of material. Only a small proportion of charges need to be transferred to create observable effects. Suppose charge is driven from one plate to another 1cm away until a potential difference of 100V has built up across the plates. Only one hundredth of a millionth of a millionth of the free electrons in the metal will have been transferred.

Friction can separate small quantities of electric charge. By rubbing two materials together electrons may be transferred from the atoms of one material to those of the other, if the second material has a greater affinity for electrons than the first. The ancient Greeks knew that a piece of amber rubbed with fur would attract small pieces of hair. The word electricity derives from the Greek name for amber—elektron.

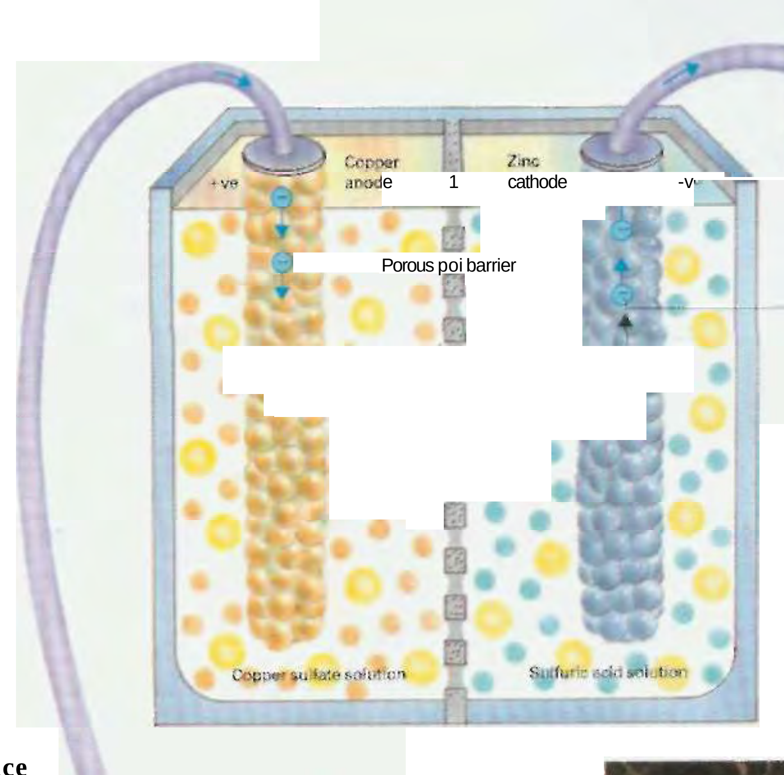
Spectacular results of the transfer of electric charge occur in lightning. Here, an electrically-charged cloud suddenly discharges, generating a current which passes through the air, raising it to very high temperatures. This causes both the flash of lightning and a shock wave of pressure changes in the air, producing the sound of thunder. The energy used in charging up the cloud initially is drawn from the atmosphere.



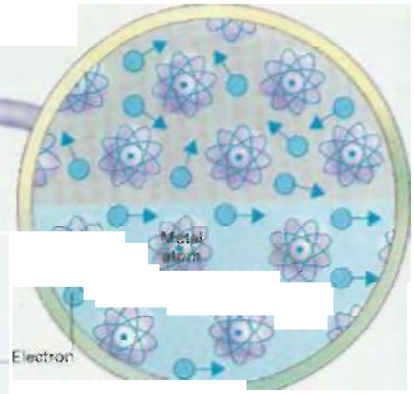
A *The Wimshurst machine generates static electricity as two insulating disks rotate in opposite directions. Charges of opposite sign are collected by metal combs at each side. When sufficient charge has built up, the electric field between the metal spheres at the top causes a spark to fly.*

•4 Lightning brightens the sky over Tucson, Arizona.

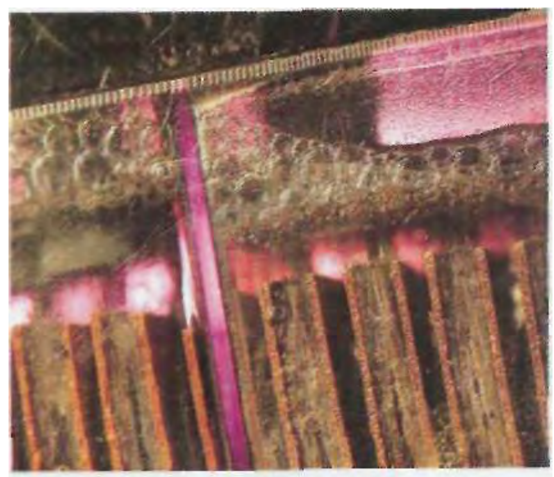
An electric battery resembles a water cistern, producing a voltage rather than a head of pressure



• *The so-called Daniell cell or battery contains sulfuric acid, the molecules of which divide into hydrogen and sulfate ions when dissolved. If the anode and cathode are connected via a circuit, a current flows as the atoms in the zinc cathode tend to give up electrons, thus becoming positive ions and joining with the sulfate ions. The hydrogen ions are attracted to the anode, where they attract electrons from the copper to form hydrogen molecules. The result is a net flow of electrons from cathode to anode. The flow continues until the cathode has been eaten away completely, or a film of non-conducting hydrogen bubbles on the anode causes internal resistance to build up.*



A *In a conducting wire, some electrons can move from .atom to atom. With no electric field, they move at random. If a field is applied, as in a circuit including a battery, the motion of the electrons is coordinated as they move towards the positive terminal.*



• *A lead-acid accumulator (as in an automobile) during recharging.*

Current and resistance

Electric potential difference is like a pressure difference in liquids. A battery behaves like a cistern, producing a voltage rather than a head of pressure. The voltage produced by a dynamo or electricity generator (^ page 58), on the other hand, resembles the pressure produced by a pump. If a battery is connected up so that it completes an electric circuit — in a bicycle lamp, for instance - the battery supplies a voltage (an electric potential difference) between the two ends of the circuit to which it is attached, or sets up an electric field in the rest of the circuit. This field makes a current flow and lights the lamp as resistance to the flow in the bulb causes the filament to glow.

Electric currents usually arise from the flow of mobile electrons. To generate a current, both a "voltage source", to provide an electric field to direct the electrons, and a "conductor", in which the electrons can move freely, are required. In metals, electrons bound loosely on the periphery of atoms wander easily from atom to atom. A small electric field will organize their normally chaotic wanderings into a steady flow in one direction.

The basic unit of electric charge is called the coulomb (C), and it is equal to the total charge of $6\text{-}25 \times 10^{18}$ electrons. The unit of current is named the ampere, or amp, for the French professor of mathematics, Andre Ampere (1775-1836), who did important work on the magnetic effects of electric currents. One ampere corresponds to one coulomb of charge flowing each second across a cross sectional slice through a wire.

Electrons flowing through a conductor, such as a piece of copper wire, are not unhindered. They can collide with the atoms of the metal and with impurity atoms. In these collisions, the electrons lose some of their energy, transferring it to the chaotic motion of the atoms. This raises the temperature of the metal, since higher temperatures are linked to faster atomic movements *(i* page 28). In ordinary filament lights, a coil of very fine wire is kept deliberately at a high temperature within a vacuum by an electric current, and the wire glows bright yellow. In an electric fire, a coil of thicker wire offers less hindrance to the current; the wire is thereby maintained at a lower temperature, and it glows red.

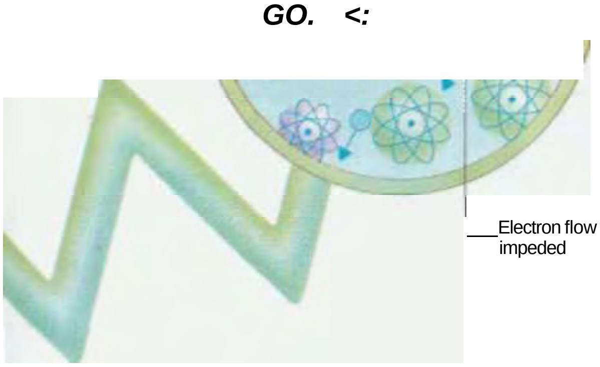
How a battery works

The basic operation of a battery is to store chemical energy and convert it to electrical energy as required. It utilizes the natural random movement in a liquid of free ions - atoms with either too few or too many electrons. The familiar kind of battery, used for example in a torch, is a "dry" battery, its liquid components being in the form of a paste. It consists typically of a carbon rod-the positive terminal-surrounded by a muslin bag containing a mixture of manganese dioxide and powdered carbon; around this, and in contact with the zinc cannister, is a layer of ammonium chloride paste.

When the terminals of the battery are connected via a circuit, electrons flow from the zinc container to the carbon rod: an electric current flows from the positive (carbon) terminal to the negative (zinc).

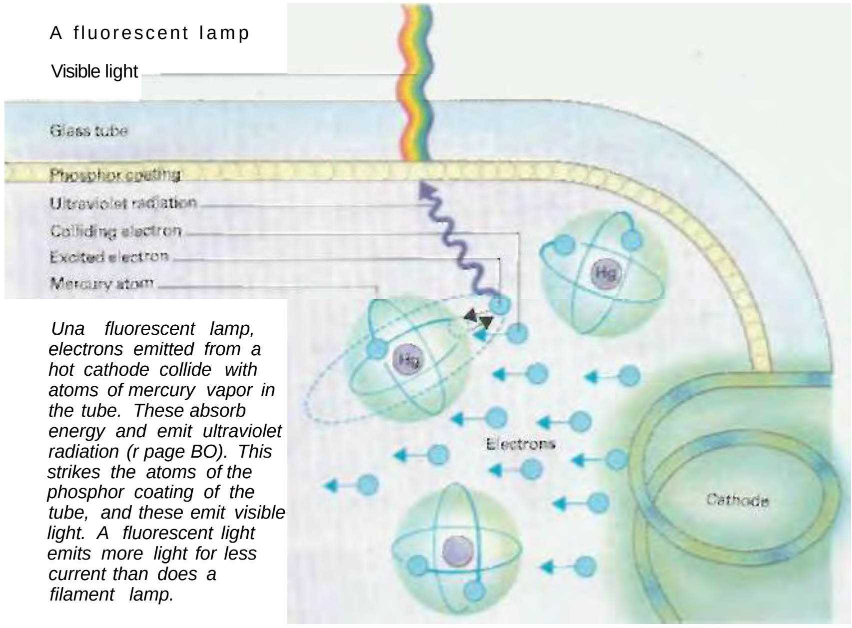
The electrons are liberated when negative chloride ions (with one additional electron each) in the paste make contact with the zinc, and form zinc chloride. The positive ammonium ions (each with an electron missing) move from the paste through the bag to the carbon terminal, where they pick up electrons and form hydrogen and ammonia. The manganese oxide prevents an insulating layer of hydrogen bubbles from building up around the carbon, it reacts with the hydrogen to produce manganese trioxide and water. The net effect is that electrons flow from the zinc to the carbon, while the ammonium chloride is decomposed.

This type of dry cell is known as a primary cell: the chemical changes cannot be reversed. In a secondary cell, the changes are reversible. One familiar secondary cell is the lead-acid accumulator used as a car battery. This consists of six cells connected in series. Nickel-cadmium "rechargeable" batteries are dry secondary cells.



+ An ordinary light bulb (here shown broken but still working) glows as the electric current encounters high resistance in the fine tungsten filament, and heats to more than 2,000'C.

A Resistance occurs when the flow of electrons is impeded by collisions with the metal atoms or with impurity atoms. The electrons lose energy to the atoms.



Una fluorescent lamp, electrons emitted from a hot cathode collide with atoms of mercury vapor in the tube. These absorb energy and emit ultraviolet radiation (r page BO). This strikes the atoms of the phosphor coating of the tube, and these emit visible light. A fluorescent light emits more light for less current than does a filament lamp.

• The ability of an object to conduct an electric current is related to its "electrical resistance": the higher the resistance, the more difficult it is for a current to flow. The resistance depends on the nature of the material in two fundamental ways. The more electrons that are free to drift in a given volume, the lower the resistance; the more obstacles in the same volume, the higher the resistance. Electrical conductors have low resistance whereas good electrical insulators have very high resistance.

Electrical resistance is a form of friction (| page 14) and causes the moving charges in a current to lose energy irreversibly. The energy lost appears as heat in the conductor carrying the current, and for the current to be maintained, the energy loss must be made good. A battery not only sets up an electric field so that current can flow through a lamp; it also replenishes the energy radiated as heat and light.

The unit of electrical resistance is called the ohm, after the German physicist Georg Simon Ohm (1787-1854). Ohm performed a series of experiments to study the flow of electricity, measuring the voltages and currents for wires of different lengths and thicknesses. The result was his celebrated law, according to which the resistance of a piece of wire, say, is given by the potential difference required to sustain a current, divided by that current. Ohm's law reflects some basic features of electrical resistance in wires and similar conductors. For example, the current through a wire is the amount of charge that crosses a cross-section through the wire each second. The larger the area of the wire, the greater the current that can be carried, and the lower the resistance. Moreover, longer wires have higher resistance than shorter ones. Indeed, wires carrying electric current are like pipes carrying water. Large currents of water can gush through short, fat pipes, but only small currents can travel along long, narrow pipes.

Ohm's law applies only to materials in which the resistance does not depend significantly on the current or the voltage applied. However, even in such materials the resistance depends on temperature, because the atoms jiggle more at higher temperatures, and this increases the number of obstacles that the electrons encounter. Semiconductor devices and conducting gases behave very differently. This is because the number of mobile electrons depends on the voltage supplied.

A light-emitting diode and a photoelectric cell are essentially similar devices that work in oppo$ite directions

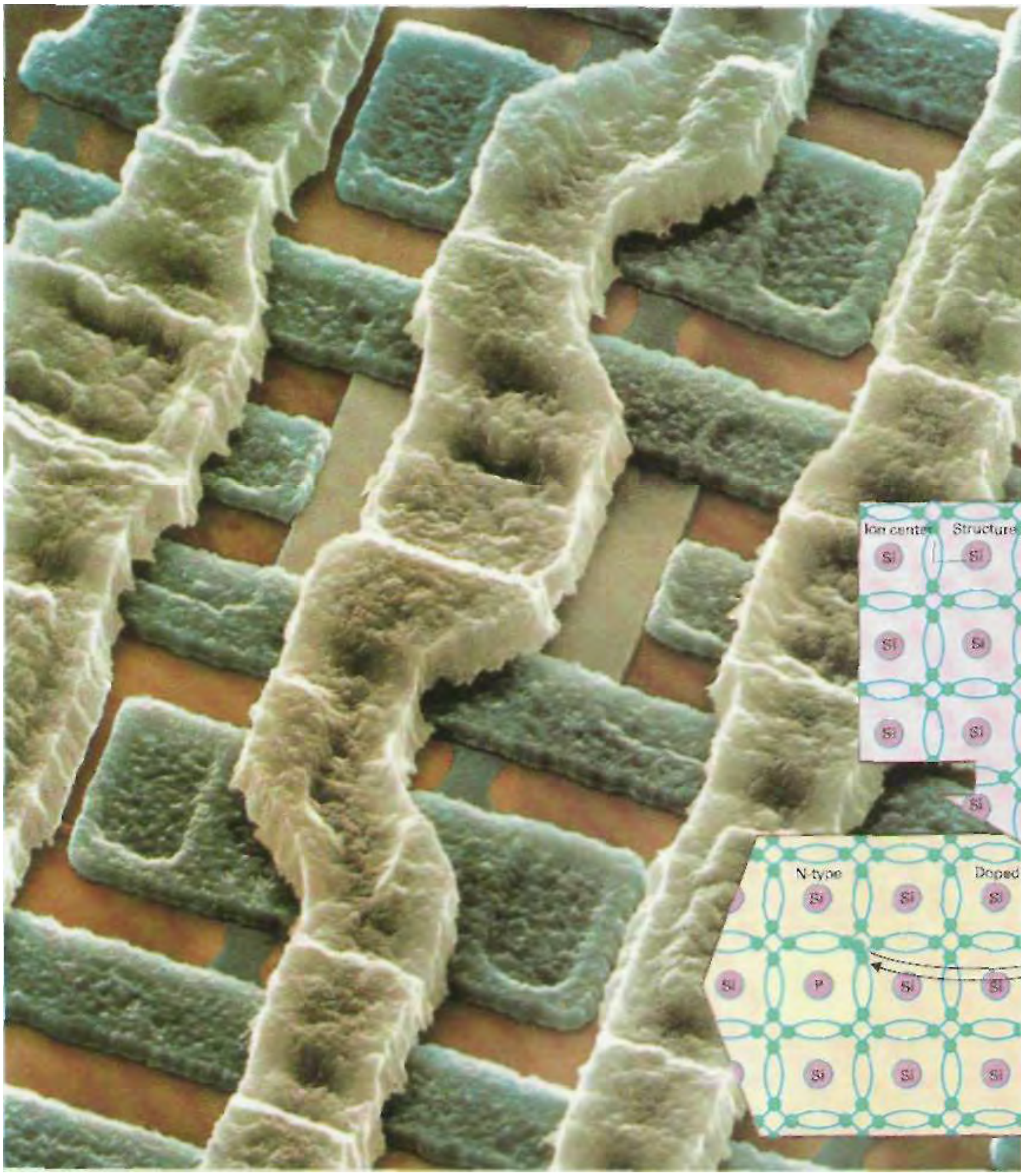
Capacitors - storing electricity

A simple capacitor consists of two parallel metal plates. It can be charged up by connecting it to a voltage source, such as a battery. The potential difference draws electrons along the connecting wires from one plate and supplies them to the other, making one plate positive and the other negative, until the potential difference between the plates balances that of the source. Then the flow of charge stops, provided the voltage applied to the capacitor is constant. A capacitor will not pass a steady current (| page 64). The source supplies energy in charging the capacitor, but this energy can be recovered.

If the two plates are now connected via a wire, with no voltage source, the potential difference across the plates drives a current through the wire and the charges on the plates fall to zero as electrons flow from the negatively charged plate to the positive. The energy that the capacitor stored is dissipated as heat in the wire connecting the plates. Thus, the capacitor acts like a temporary voltage source.

The field in a capacitor

If the two parallel metal plates in a capacitor receive equal amounts of opposite charge the charge spreads across the plates to give a constant electric field, proportional to the charges on each plate divided by the area of each plate. The electric potential of the plate carrying positive charge relative to the negative plate — in other words, the voltage across the capacitor-is given by the field multiplied by the distance between the plates. Capacitance is defined as the amount of charge a capacitor can hold divided by the potential difference that the charge produces. If a capacitor holds a charge of one coulomb on each plate, and has a potential difference of one volt between the plates, then its capacitance is one farad (F), the unit of capacitance, which is named in honor of Faraday. In practice, capacitances as high as one farad or so are rare.



•< T *Semiconductor memories are thousands of memory cells constructed on a single silicon chip. Each cell can trap charge with electric fields set up both by external voltages and the internal movement of charge between different types of silicon. Silicon atoms are bound together by covalent bonds between the four outer electrons. Doping with atoms of one electron more, or less, creates n-type orp-type silicon. Electrons tend to diffuse one way across a junction between the two types and "holes" the opposite way. This leaves the n-type side positive, the p-type side negative, and so establishes an internal field.*

Using capacitors

Capacitors are used for a variety of purposes both to store electrical charge and energy and to control the response of electrical circuits. For example, a camera with an electronic flash unit contains a capacitor to store the energy for the flash. Moreover, capacitances as big as one farad are nowadays used in place of batteries to back up semiconductor computer memory during power failures. And because capacitors can store charge they can be used as tiny "dynamic" memory elements in microchips.

Capacitors also have important applications in conjunction with other circuit elements. In particular, one of their key functions is to control varying electric currents (} page 64). They are also used, for instance, in "filters" to remove unwanted signals in sensitive pieces of equipment such as television sets and audio amplifiers.

Semiconductors and silicon chips

In a metal the outermost one or two electrons in each atom are free to move through the material underthe influence of an electrical force. In an insulator none of the electrons are free to move in this way. There is also a third class of materials, which are insulating at low temperatures, but which become conducting as the temperature rises. At higher temperatures, the electrons are more energetic, and some become free to drift readily through the material, just as in a metal. Such materials are "semiconductors". Understanding conduction in semiconductors and indeed in conductors, has been possible only by applying quantum mechanics to the solid state.

One way to provide a semiconductor with mobile electrons is to add,or "dope", small quantities of another element-an "impurity" - in which the atoms contain more outer electrons. It is also possible to add atoms that contain fewer outer electrons than the basic semiconductor. In this case, the impurity atoms capture electrons from the semiconductor, leaving "holes" among the outer electrons in the atoms of the semiconductor. These holes simulate positive charges andean thus drift freely under the influence of an electric field, just as the negatively-charged electrons do.

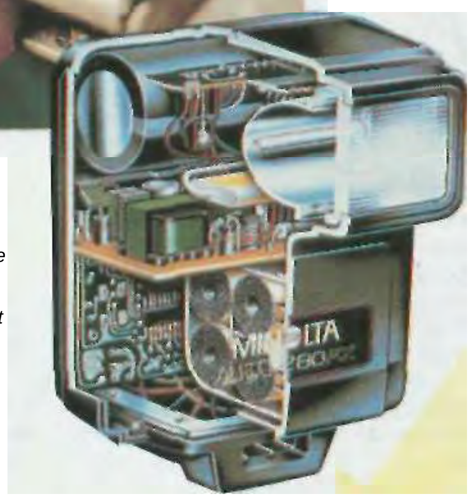
Electronic engineers can accurately control the electrical properties of a semiconductor by doping it with impurities that donate extra electrons, or which mop up electrons and thereby provide holes. Further control is possible through the appropriate application of voltages. Thus in a transistor, the current flowing through the semiconductor can be altered or even switched off by suitably applying a voltage that is different from the one maintaining the flow of current. This effect allows transistors to operate as switches and as amplifiers.

Silicon (Si) is an important semiconductor, which can be grown into large single crystals - as big as 10cm across. A thin slice of such a crystal is the basis of the silicon chip. Impurity atoms can be diffused into the surface of the silicon wafer to make a very large number of tiny circuit elements such as transistors or capacitors, on the same chip. With appropriate connections, vast amounts of information can be stored and processed.

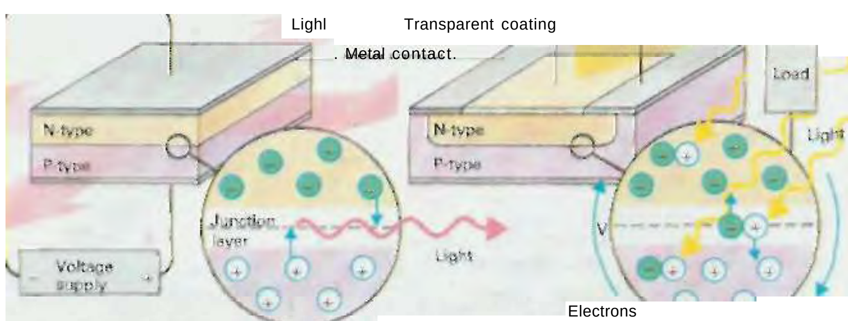
Both the voltage and the current fall to zero as the charge on the plates decreases, rapidly at first and then more slowly. The time taken for the capacitor to charge up and to discharge is inversely proportional to the resistance and to the value of the device's capacitance.

The capacitance can be increased by filling the space between the plates of a capacitor with a "dielectric". This is a material that does not conduct much electricity, but can be electrically polarized - the molecules within it can be pulled apart a little by an electric field. Inside a charged capacitor, the negative charges in a dielectric are pulled towards the positively-charged metal plate, and the positive to the negatively-charged plate. Thus, a layer of negative charge forms near the positive plate and a layer of positive charge near the negative plate. Both the electric field between the plates is reduced and the potential difference between them. The plates' charge is the same, however, so the capacitance - charge divided by potential difference - increases. Commercial capacitors contain dielectrics to give higher capacitances.

•< T *Capacitors have a range of uses, which depend on their ability to store charge or on their response to alternating currents. As charge storage devices, they are used as cells in semiconductor memories (opposite). They also store charge in electronic flash guns in cameras (below). Variable capacitors (left) were once used in tuning circuits in radios, although variable inductors are now used.*



T *An external voltage applied across a junction between p-type and n-type semiconductors can cause electrons and holes to move into the junction layer and recombine, to give energy. This energy appears as light in certain semiconductors and is how light emitting diodes work. In a solar cell, absorbed light creates electron-hole pairs. At the junction, the internal field separates these charges and a current flows.*



See also
Magnetism 46-8
Electromagnetism 57-64
Using the Elements 73-8

Electrolysis

When an electric current is passed through a solution of ionic solid such as a metal salt, the different ions in the solution migrate towards either the positive electrode (called the anode) or towards the negative electrode (called the cathode). Negatively-charged ions move towards the anode and positively-charged ions move towards the cathode. When ions reach the electrodes they either take up electrons, if they have an overall positive charge, or give up electrons, if they have an overall negative charge, to form atoms or molecules. This phenomenon is known as electrolysis. It is used in industrial chemistry to form compounds such as sodium hydroxide, and in electroplating to build up a coating of layer upon layer of metal atoms.

Under the influence of the electric current, ions in the solution move towards the anode and cathode. The positively charged metal ions travel to the cathode and are deposited on the object to be plated. The anode degrades to replenish the electrolyte with metal ions.

Faraday and electrolysis

The knowledge of electrolysis stems from the experiments performed by the British physicist and chemist Michael Faraday (1791-1867) in the first half of the 19th century.

Faraday determined the quantitative relationship between the amount of electricity passed through an electrolytic cell and the chemical change it produces. He showed that the weight of a substance converted by electrolysis is proportional to the quantity of electricity used. One mole of a substance that gains or loses one electron during electrolysis, for example, is deposited by 96,485 coulombs, or ampere seconds, of electricity. In general, 96,485 times n coulombs of electricity are needed to liberate one mole of substance - where n is the number of electrons lost or gained by each atom or molecule during the reaction. The quantity 96,485 coulombs is known as a Faraday but in practical electroplating the more important unit is the ampere-hour, a unit equivalent to 3,600 coulombs. One Faraday equals 26-8 ampere-hours.

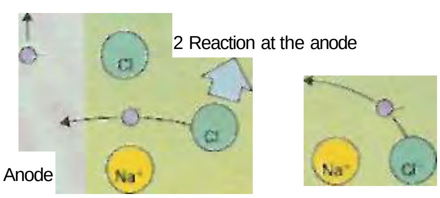
Coating metals

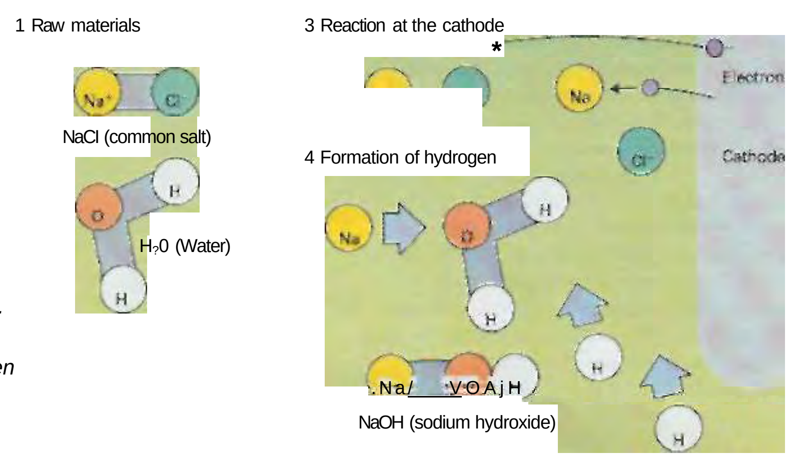
Many metal objects and tools can be coated with a thin layer of another metal to make them more resistant to wear and attack by corrosion. These coatings can be applied by dipping in a bath of molten harder or more resistant metal, or by using an electrolytic technique known as electroplating. Electroplating can give an extremely thin and uniform coating to a host of metals.

Objects to be electroplated may be suspended from the cathode of an electrolytic cell in a bath or vat containing the solution of a metal salt— the electrolyte. An anode, often made from the metal to be electroplated, is also suspended in the vat and a low DC electric current is passed through.

All sorts of metal objects are electroplated, from nuts and bolts through to tablewear and parts of automobiles. The metal fenders of automobiles are usually made of steel, and then electroplated with nickel and chromium. The nickel and chromium not only help to make the steel more resistant to corrosion, but they also give a shiny finish which is used as decoration. Silver-plated tableware is common. The silver coating protects the base metal of the tableware from attack.

Many metals can be plated using simple plating baths. Often the process is continuous and an integral part of the manufacture of the metal object or tool. Strips of steel are electroplated with tin to provide the tin plate from which "tin" cans for food are made; some steel drills are given extremely thin plates of nickel and chromium to make them harder. Metal alloys can be electroplated if the metal salt solutions are mixed in the electroplating bath. Plated brass made in this way is indistinguishable from cast brass.





Electrolysis is the separation of chemical species by passing a current through them, following a process that resembles that of an electric cell or battery. If a current is passed through a solution of common salt (NaCl) in water (H_2O), the negative Cl ion is attracted to the anode, where it gives up its excess electron and is given off as chlorine gas. Similarly, the positive Na ion is attracted to the cathode, where it picks up an electron and bonds with the water molecule to form sodium hydroxide (NaOH), while hydrogen gas is given off.

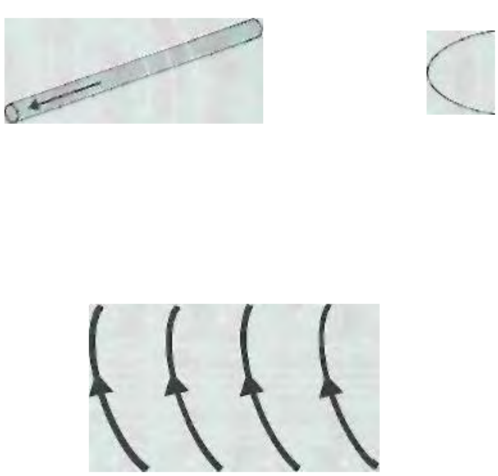
Electromagnetism

8

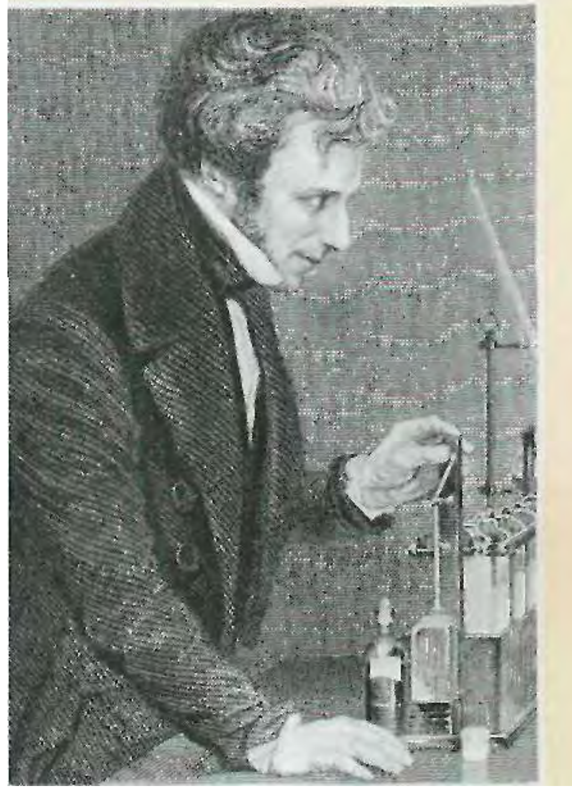
Magnets, currents and fields... The electromagnetic spectrum...Alternating current... Transformers... PERSPECTIVE...Michael Faraday...Motors and dynamos-Light and the electromagnetic spectrum... The transmission of power and signals...Inductors and capacitors

Electric charge in motion - electric current - is the origin of magnetic fields (4 page 45). Indeed, magnetic fields are not truly distinct from electric fields; rather, both are aspects of a general "electromagnetic field". In 1820, the Danish physicist Hans Christian Oersted (1777-1851) discovered that a current flowing in a wire deflected a compass needle nearby. Before the end of the year, Ampere had extended Oersted's work and concluded that all magnetism is due to tiny electric currents. The form of the magnetic field generated by a tiny piece of wire carrying an electric current is complicated. The direction of the field is at right angles to the wire, and at right angles to the line joining the wire to the point in space at which the field is measured. The strength of the field is proportional to the current and to the length of the wire, but it also depends on distance according to the inverse-square law, and on the angles between the wire and the joining line. The strength of the field is greatest at right angles to the wire. To calculate the field due to the current in a specific wire, such as a loop or a coil, means adding all the contributions from tiny portions of the wire, but in some instances the answer has a relatively simple form. For example, the field due to a solenoid - a coil with many turns - resembles that of a bar magnet with a north pole at one end and a south pole at the other (| page 45). Indeed, the magnetic fields produced by permanent iron magnets are due to electric currents, not from coils of wire but from electrons spinning on a microscopic scale.

Electricity and magnetism



A *Electric current flowing along a wire creates a magnetic field encircling the wire (A). The direction of the field accords with the rotation ofa right hand screw moving the same way as the current flow. In a wire bent into a single turn, the magnetic field consists of many loops (B), while in a solenoid (C) the effects of many turns of wire add together to give a net field like that ofa bar magnet.*



A *Faraday Is known as one of the most outstanding experimenters of all time. He is remembered for his work on electromagnetism, static electricity and chemistry.*

Michael Faraday

Michael Faraday's life story is the classic tale ofa great scientist's rise from humble origins. The son ofa blacksmith, born in London in 1791, he became a bookbinder's apprentice at the age of 14. However, in 1813 he was taken on as assistant to the chemist Humphry Davy (1778-1829) at the Royal Institution in London. By 1825, Faraday had become Director of the Laboratory there.

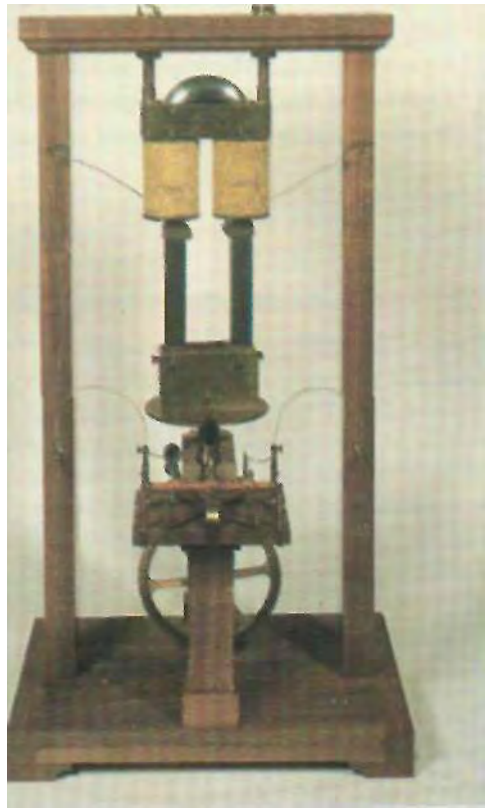
Faraday was a prolific experimenter, and his name is attached to many phenomena both in physics and chemistry. Though no mathematician, his ideas on electric and magnetic forces gave birth to the theory of electromagnetic fields. Following the work on magnetism by the English scientist William Gilbert (1540-1603) (fpage 47), Faraday conceived the idea of electric as well as magnetic "lines" of force, to depict the fields around electrified and magnetized objects. He was one of the first scientists to recognize the reality of these fields, and his work was later of great value to the Scottish physicist James Clerk Maxwell (1831-1879) in his synthesis of electricity and magnetism O page 60).

Faraday's work ranged from the study of capacitance, to the chemical effects of electric currents in the process of electrolysis, to the discovery of benzene. His most important discovery is arguably that of electromagnetic induction. In 1831, he found that if a loop of wire moves through a magnetic field, then an electric current is induced in the wire. This is the principle that underlies the production of electricity.

The interaction of electrical and magnetic fields

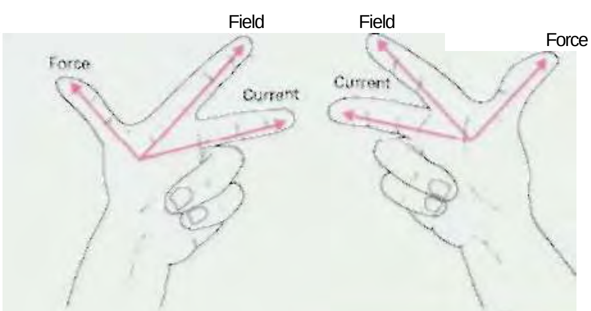
The action of a magnetic field on an electrically-charged particle is wholly different from that of an electric field. The force exerted on the charged particle is proportional to its charge, to the velocity of the particle, and to the part of the field that is at right angles to the particle's motion. The force acts at right angles to both the magnetic field and the motion of the particle. Such forces can move matter in bulk, by acting on the electrons flowing in a current. This force underlies the electric motor, in which a coil of wire carrying a current is influenced by the field between the poles of a magnet.

The reverse effect is put to good use in the electric generator (| page 62), where mechanical energy is turned to electrical. The electric fields described earlier are generated by electric charges, and the magnetic fields by charge in motion - in other words by electric current. Electric fields are also generated by magnetic fields that change with time. The electric fields produced in this way are at right angles to the magnetic fields. In a similar way, magnetic fields are not only generated by electric currents, but also by electric fields that change with time. The generated magnetic fields are at right angles to the changing electric fields, and again the magnetic field lines form closed loops about (he changing electric field.

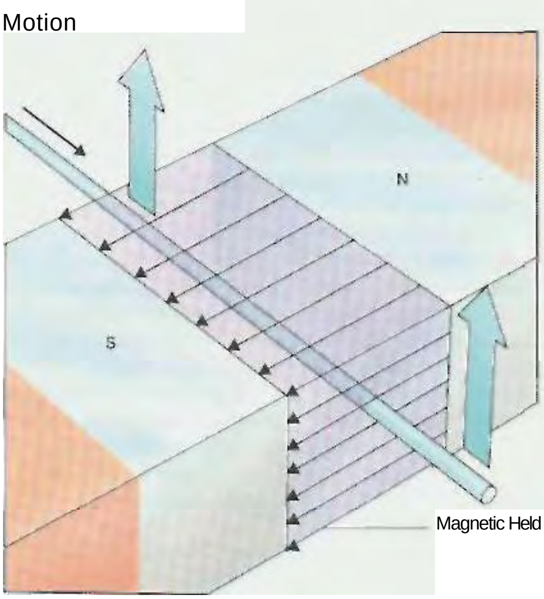


• A reconstruction of one of the earliest devices to study the connection between electrical and magnetic fields, built by the physicist H. Pixiiin 1821.

• Two rules, based on the directions of the thumb and first two fingers pointing at right angles, describe the directions of current flow, magnetic field and motion in motors and generators. The left-hand rule pertains to motors, the right-hand rule to generators. The forefinger points to the field, the thumb to the motion, and the middle finger to the current.

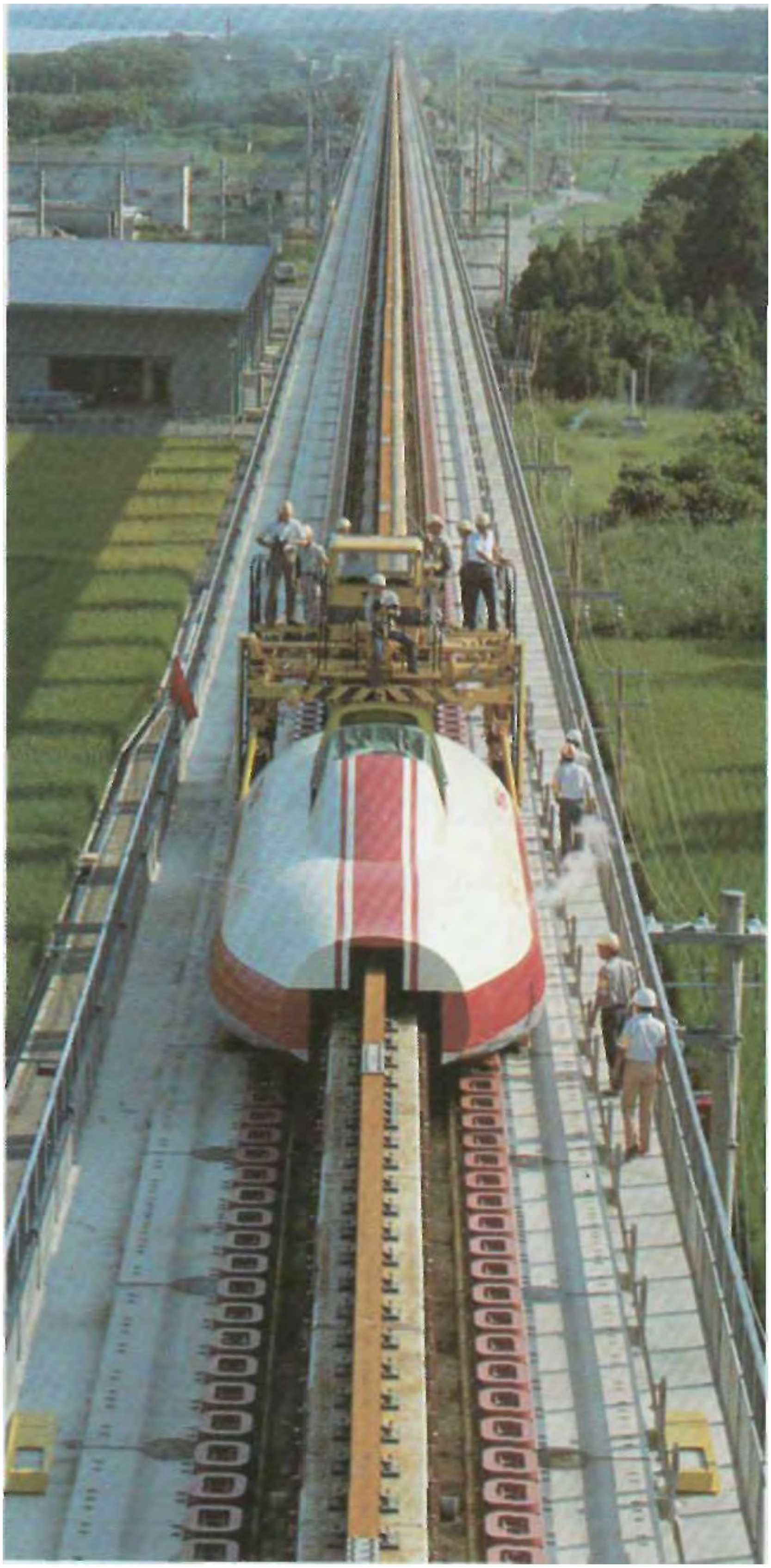


•• A magnetic-levitation Imaglev) train is an application of the electric induction motor, producing a linear rather than a rotational motion.



A *When a wire is moved in a magnetic field, a current flows if the wire is part of a circuit, because the electrons in the wire experience a force. This is the principle that underlies the operation of generators.*

• Many children's toys incorporate simple direct-current electric motors operating a low voltage, whether from batteries or adapted from the mains supply by means of a transformer I $ page 63).



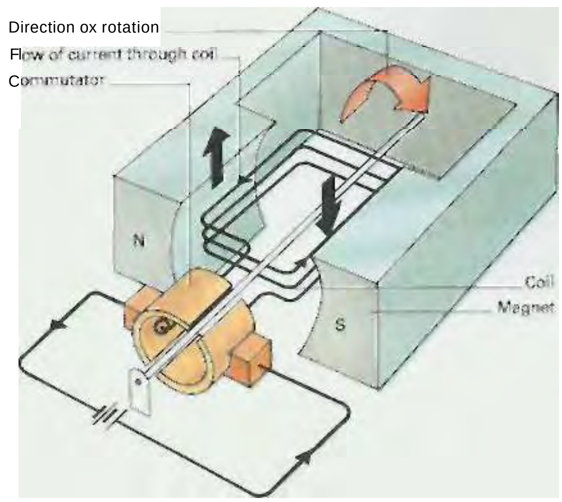
Dynamos and motors

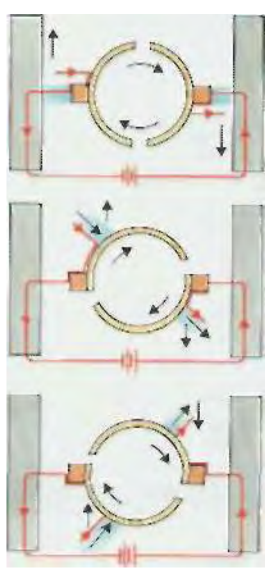
Generators (dynamos) and motors are related in the way they work. A dynamo converts energy of motion to electrical energy. An electric motor converts electricity to motion. The common factor between the two pieces of equipment is the effect of a magnetic field on moving electrons.

The electrons in a wire moving through a magnetic field feel a force as they travel through the field. The force sets the electrons in motion along the wire. This is how a dynamo works. The force on the electrons creates an electrical pressure like the water pressure generated by a pump.

In electric motors, a current is set up through a wire held in a magnetic field. The moving electrons then feel a force due to the magnetic field, which makes the wire itself move. If the wire is in the form of a loop, the forces acting on the two sides of the loop make it spin in the magnetic field, andean keep it spinning as long as the current continues.

A simple motor requires an alternating current to work -as the coil turns between the poles, so the current is reversed to ensure continued rotation. For direct current, a commutator reverses the current through the coil of each half rotation.

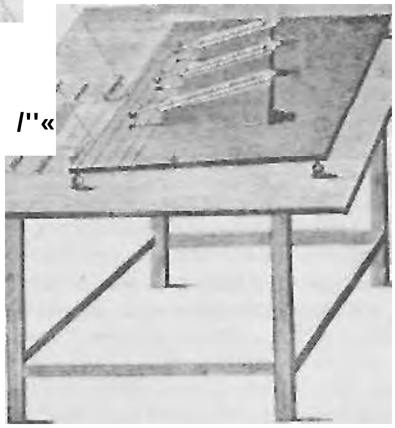




•4 A *In a simple direct current motor, current passes from a battery through one or more loops of wire held between the poles of a magnet. The magnetic field exerts a force on the electrons carrying . the current and this makes the coil begin to rotate - as the force is in opposite directions on opposite sides of the coil. If the curren t remained in the same direction, the forces on the coil would make it rotate in the opposite direction once it was past the vertical position between the poles. To keep the coil rotating, the current must reverse direction between half-cycles. This is achieved by an arrangement of a commutator—a split ring with carbon brushes on either side as contacts.*

The Electromagnetic Spectrum

/''«



A • *In 1800, the British astronomer William Herschel (1738-1822), best known for discovering the planet Uranus, made another important discovery. He was studying the heating effect of the Sun's radiation, using a thermometer with a blackened bulb, when he found that beyond the red end of the visible spectrum, the thermometer continued to absorb heat and show an increased temperature. This was the first evidence for infrared radiation - "beyond" the red - with wavelengths longer than visible light. Infrared radiation is emitted by all hot objects, from stars to electric fires, and infrared imaging provides a way of "seeing" at night.*

Electromagnetic radiation

The fact that magnetic fields feed on changing electric fields and vice versa means that electromagnetic "waves" can propagate through space. The electric and magnetic fields in these waves are at right angles to each other and to the direction in which the wave is traveling. Such electromagnetic waves are familiar as light, radio-waves and X-rays. The only difference between these phenomena is the wavelength of the waves. The source of all such waves is the movement of electrons. Stellar radiation, generated by electrons moving randomly around in the hot gases at the stellar surface, can travel many light years across empty space. In a broadcasting station, however, the correlated motion of electrons in the trans-mitting antenna produces the radio waves.

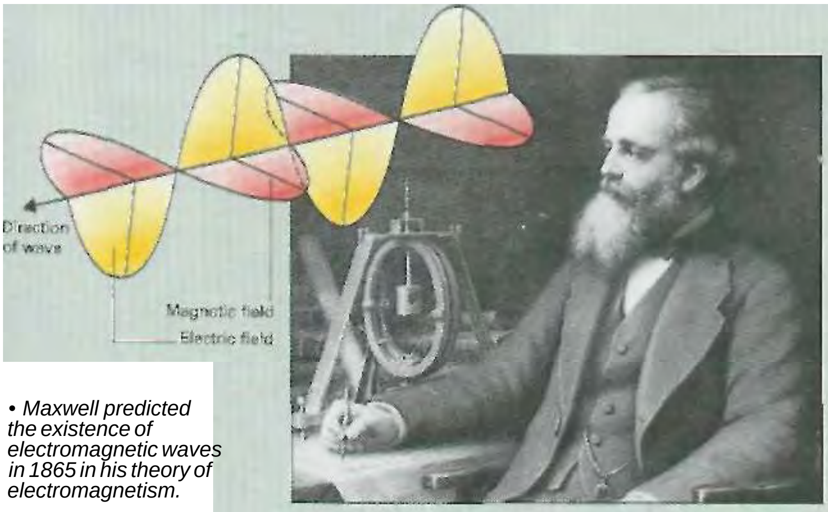
>• *Electromagnetic waves consist of mutually perpendicular electric and magnetic fields that oscillate at the same frequency and propagate in the same direction at a velocity of about 300,000 km/s - the speed of light- in a vacuum. They are transverse, unlike the longitudinal sound waves Hpage22): the electric and magnetic fields oscillate at right angles to the direction of propagation. The waves arise as a result of the relationship between electric and magnetic fields, changes in one giving rise to changes in the other.*

f *Electromagnetic waves have a huge range of possible wavelengths or frequencies, giving rise* to *an electromagnetic spectrum that covers more than 19 decades in wavelength and frequency. The spectrum stretches from gamma rays at the shortest wavelengths through X-rays and visible light to microwaves and radiowaves. Shown below is the Orion Nebula seen with radiation in different regions of the spectrum. from the X-ray image on the left, through the optical and infrared pictures, to a radiowave image (right).*

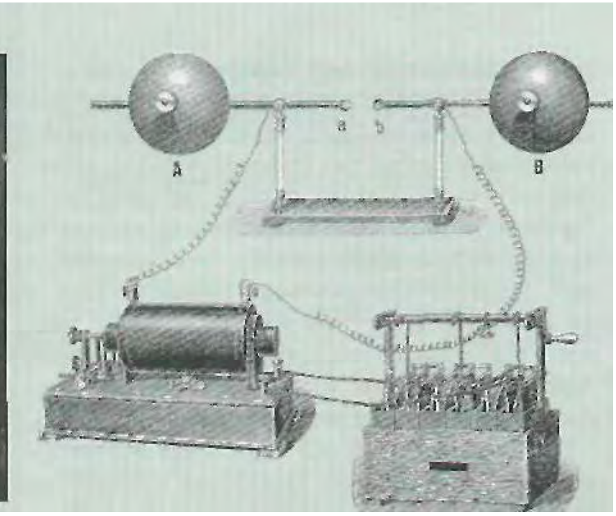
Wavelength *Im)* 10 10- io: 10

Gamma rays X-rays UUra-violet radiation

Frequency *(Hz)* 10" 10" 10" *w* —r 10" 10" 10'6



• *Maxwell predicted the existence of electromagnetic waves in 1865 in his theory of electromagnetism.*



The discovery of electromagnetism

In 1865, the Scottish physicist James Clerk Maxwell (1831-1879) brought together all the known laws of electricity and magnetism and reduced them to four basic equations. He thus revealed a symmetry between electrical and magnetic phenomena, and also introduced a new ingredient. It was already known that a changing magnetic field generates an electric field and that an electric current in a wire creates a magnetic field. Maxwell's extra claim was that a changing electric field can also produce a magnetic field. Including this effect meant that the equations can be seen to describe a kind of wave.

Wave motion is familiar as the ripples on water, but in the case of Maxwell's waves nothing is moving materially. The equations describe the changing wavelike pattern of fields across space. An electrical or magnetic "disturbance" can set off such a pattern of changing fields. Maxwell found that the speed of propagation must be close to the velocity of light, and so he proposed that the two speeds are equal: in other words, light is a form of electromagnetic wave. In taking this step, Maxwell acknowledged the influence of the English scientist, Michael Faraday (1791-1867), who 20 years earlier had considered light as an electrical phenomenon.

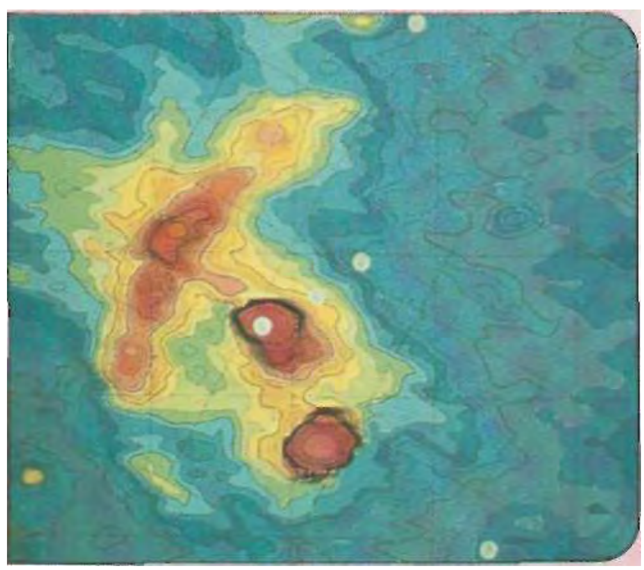
Maxwell's idea of light as an electromagnetic wave was not fully accepted until 1889. The German physicist Heinrich Hertz (1857-1894) demonstrated that electric currents could generate "signals" — radio waves — that move at the speed of light. These waves exhibited the phenomena of reflection, refraction, diffraction and interference, that the wave theory of light had already explained.

A *Maxwell's theory of electromagnetic waves was proved experimentally in the late 1880s by Heinrich Hertz. Hertz connected the terminals of an induction coil (lower left) so as to charge up a capacitor. Sparks eventually passed between the two small balls, discharging the capacitor, but in an oscillatory fashion due to the presence of the inductor in the same circuit. This oscillating current produced olectromagnotic waves which could induce a spark between small spheres in an almost closed loop of wire some distance away.*

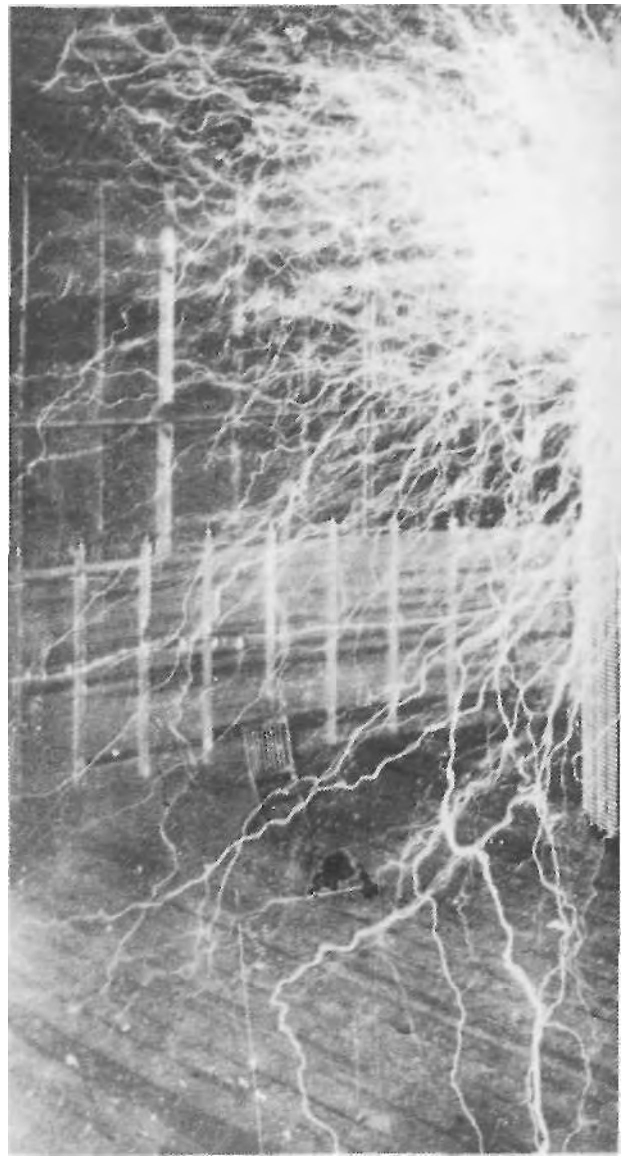
10° 10' 10'

Visible light Thermal radiation Microwaves Radiowaves

10>- 10" 10'3 10-' 10" 10"' — r 10' 10°



A message moves down a telephone wire at the speed of light, although the electrons in the wire travel at only 0.1 millimeters per second



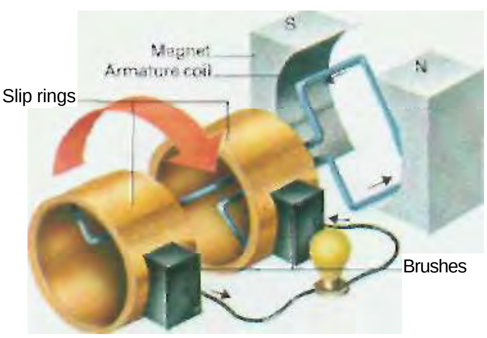
Alternating current and transformers

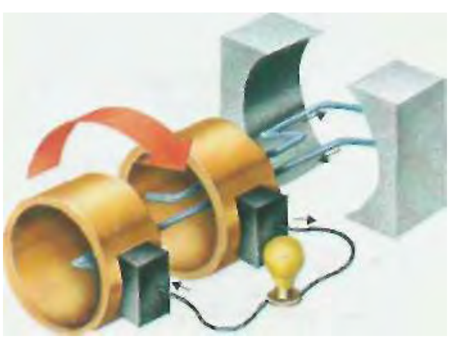
The electricity supplied to most homes and factories is in the form of an "alternating voltage", which changes regularly between a positive value and an equal but opposite negative value with respect to a zero defined by the Earth's natural potential (referred to as "earth" or "ground"). Thus, whereas in direct current the electrons pass in a single direction down the wire, in alternating current they move back and forth as the potential changes. This alternating voltage is generated at power stations by rotating generators which extract energy from high-pressure water (hydroelectric generators) or from high-pressure steam. In the latter case the steam is heated by burning coal or oil, or with heat derived from nuclear fission (*1*, page 120).

The main reason for supplying electricity with an alternating voltage is related to the problem of transmitting electrical power with minimum losses. It generally pays to transmit at much higher voltages than are practical to generate or to use, so some device is needed to "step up" the voltage prior to transmission, and to "step down" the voltage at the receiving end. The device known as the "transformer" proves to be very efficient at changing voltages from one level to another, but it works only with alternating current.

In a typical transformer, two coils of wire surround the opposite sides of a square loop of iron. An electric current flowing in one coil generates a magnetic field, which is proportional in strength to the current and to the number of turns in a coil. The iron amplifies and traps this field which then threads the second coil. If the trapped magnetic field is changing - which it is if the current in the first coil is alternating - it generates a circulating electric field, which sets up a voltage within the second coil. This voltage is proportional to the rate at which the magnetic field is changing, and to the number of turns in the second coil. If the second coil has more turns than the first, the voltage is increased; if it has fewer turns, the voltage is decreased.

The AC generator

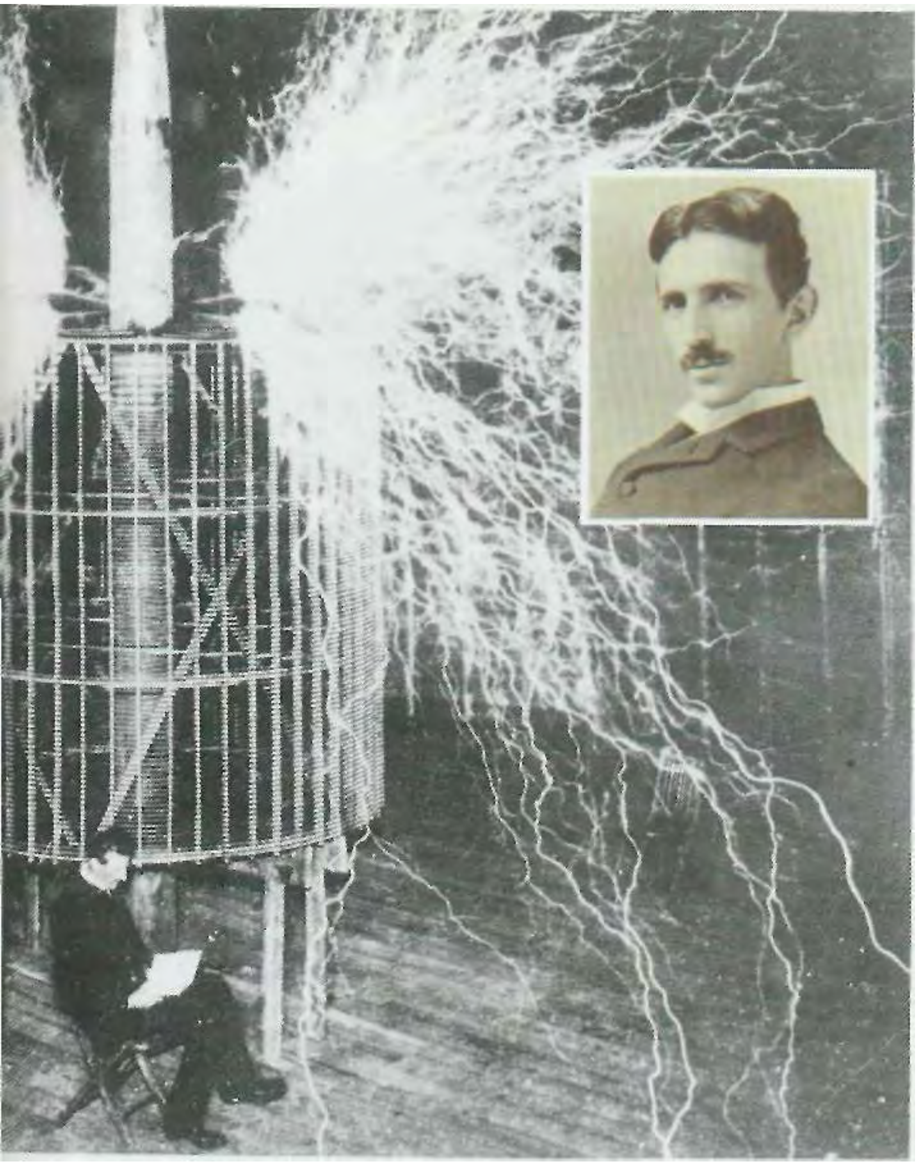




A An alternating current generator resembles an AC motor: as a coil is turned by mechanical means between the poles of a magnet, a current is generated in the coil. As the coil turns between the poles the current is reversed - if the rotation of the coil is kept constant, a steadily alternating current results.

• A simple transformer consists of two coils wound on linked iron cores. An alternating voltage across one coil (the primary) sets up an alternating current in the coil and hence induces a changing magnetic field in the iron core. This varying field induces an alternating current in the other coil {the secondary). The size of the voltage in the secondary coil depends on the number of turns in the two coils. With the same number, the voltages are equal; with more coils in the secondary, the induced current is increased (a step-up transformer), while with fewer turns the induced current is decreased.

Input voltage V
Output voltage V"
Iron core
Input voltage V
Output voltage 2V
Input voltage V
Output voltage VW



M The Croatian-born physicist Nikola Tesla 11875-1943) emigrated to the United States in 1884, where he worked with T.A. Edison. In 1888 he developed the AC induction motor, and improved the techniques of power transmission.

Transmission of signals and power

One of the most profound ways in which electromagnetism affects daily life is in communications. Electrical effects can pass information between two distant points by transmitting signals that can be interpreted as simple messages, as sounds, or even as color images together with sound, as in a television signal.

Signals can be transmitted by "modulating " electromagnetic waves (that is, by modifying them slightly) or by changing the level of current flowing along a wire. The earliest practical communication system of this kind was the telegraph, developed in 1844 by the United States inventor, Samuel Morse (1791-1872). In "Morse code", a message is translated into a series of "ons" and "offs", where there are two kinds of "on ", one lasting for a longer time (a "dash ") than the other (a "dot"). A compass needle at the end of a length of wire is sufficient to reveal if the current has been switched on or off at the other end.

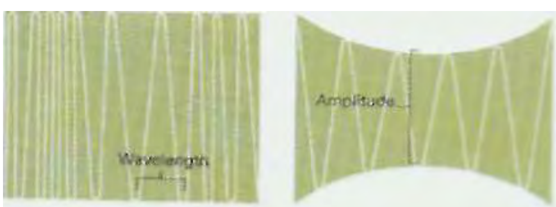
Nowadays, radio and television signals are transmitted via electromagnetic waves although ultimately they travel along wires. The signals in the electromagnetic waves induce currents in a TV or radio aerial which then propagate signals through a cable to the receiver. Local telephone calls are transmitted along wires, but international calls may involve a stage in which the signal is carried by electromagnetic waves to a relay satellite.

Electric currents can be used to carry electrical power, rather than information. Large quantities of power are distributed to homes and factories in towns and cities along wires - power cables — sometimes across distances of hundreds of kilometers. These cables have electrical resistance, which means that some energy is lost in transmitting the power. To minimize these losses, very high potential differences (voltages) are used. The power lost in overcoming resistance is proportional to the resistance multiplied by the square of the current. The transmitted power is equal to the voltage generated multiplied by the current supplied. Thus to maximize the transmitted power, while minimizing the power lost, it pays to keep the current low and the voltage as high as is practical - at 400,000V, for example. However, such high voltages are not suitable in most industrial or domestic applications. When the power lines reach their destination, the voltage is reduced to an appropriate level, with little power loss, by using transformers.

Both the signals transmitted along telephone lines and the energy transmitted along power lines travel at the speed of light, not at the speed at which the electrons drift through the wires (in the region ofO- Imm/s). The wires and the drifting electrons they contain act as guides for electromagnetic fields, which carry both signals and energy at the speed of electromagnetic waves, that is, the speed of light.

Frequency modulation

Amplitude modulation



-4 A signal is carried on an electromagnetic wave by modifying the frequency of a wave with constant amplitude IFM), or by modifying amplitude with constant frequency (AM).

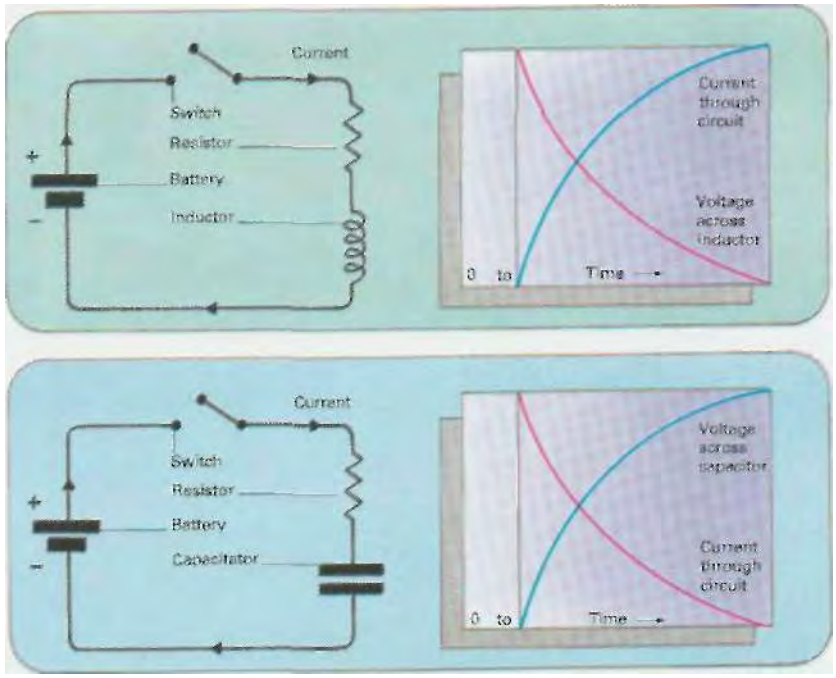
T Television signals are carried from transmitter to receiver as FM waves.

See also
Magnetism 45-8
Electricity 49-56
Using the Elements 73-8
The Quantum World 87-9B

Inductors

An electric current flowing through a coil of wire generates a magnetic field that threads the coil. If the current changes with time, the magnetic field also changes, and therefore generates an electric field. However, the force that the electric field exerts on the electrons is such that it opposes their motion and impedes the changing current.

If such a coil - known as an inductor—is connected to a constant voltage source, then the current through the coil will rise slowly to a value equal to the applied voltage divided by the resistance of the circuit. Unlike a capacitor, an inductor does not impede a steady current, although in practise the wire will have some resistance. However, the potential difference generated in the inductor by the changing current is proportional to the rate at which the current changes. Thus, again unlike a capacitor, an inductor will not pass alternating currents with very high frequencies. This contrasting behavior of inductors and capacitors means that the two devices can be used together as a "filter". Hifi equipment and other devices that seek to filter a clear signal out of its background "noise" make use of this effect.



V

In Out

Mixfid frequency output I ow pass fitter 1 ow frequency output

VYi AA*—* I1

High pass filter High frequency output

Band pass filler Medium frequency output

•< *The complementary responses of capacitors and inductors can be combined together in filter circuits which operate on alternating current. The values of capacitance and inductance can be chosen so that the circuit passes current only at low frequency, high frequency, or in a range of frequencies. Conversely, the filter can be designed to block unwanted signals of particular frequencies.*

A *Inductors and capacitors respond in opposite ways when a voltage source in the circuit is switched on. In an inductor (topi the current rises as the voltage across the coil falls. With a high frequency AC supply, the inductor "blocks" the current. In a capacitor, the current falls as the voltage rises to its full value, but at high frequencies the capacitor "passes" current.*

• *The electromagnets that guide beams of subatomic particles around accelerators require DC voltage, which can come from the grid, but only after rectification, converting AC to DC. This process still leaves "ripples" on the DC voltage, which must be reduced to less than 0-01 percent of the total voltage for the magnets to produce fields stable enough to keep the particles on track. Filters are therefore crucial.*

Atoms and Elements

9

The elements... Theories of the atom...Structure of the atom...Nuclei and electrons... Atomic weight and number...Electron shells and orbitals...The Periodic Table...PERSPECTIVE...Dalton's atomic theory... Calculating atomic mass...Predicting unknown elements...Features of the Periodic Table

Every substance in the universe, from hydrogen to the complex proteins that are essential to life, is made of either a pure element, or a number of different elements bonded together to form molecules. The distinction between molecules and elements is not an obvious one, but the identification of elemental materials, notably some metals, was achieved by the alchemists. In the 18th century the French chemist Antoine Lavoisier (1743-1794) identified the element oxygen as being responsible for combustion in air, and more elements were discovered throughout the 19th century.

The study of matter on the molecular or elemental level was greatly assisted by the revival of the atomic theory by the English chemist John Dalton (1766-1844). According to this theory, all matter is made up of atoms. The atom is one of the basic structural units of matter, and is the smallest particle of an element that can take part in chemical combination to form a compound. This concept was originally proposed by the Greek philosopher Democrilos (*c.* 400 BC), who argued that there existed minute particles of which all matter was made. These particles, he suggested, were indivisible and indestructible. The word atom is derived from the Greek word meaning "that which cannot be divided".

Today it is known that, although atoms are the smallest fragments into which an element may be divided without losing its particular properties, they are not themselves the fundamental units of matter. Rather, they have a structure and are made of more fundamental particles - protons, neutrons and electrons. At the center is a very small, dense nucleus which contains the relatively heavy protons and neutrons (^ pages 79-86). These have almost equal masses, but protons possess a positive charge, whereas neutrons have no charge at all. Overall, however, atoms are electrically neutral, since the positive charge of the protons is balanced by the negative charge of the electrons. These exist at a relatively large distance from the nucleus, and the space within which they can move is, in atomic terms, vast. We know from mathematical calculation of probability that electrons exist in volumes around the nucleus; these volumes are known as atomic orbitals (f page 68).

Thinking of the atom

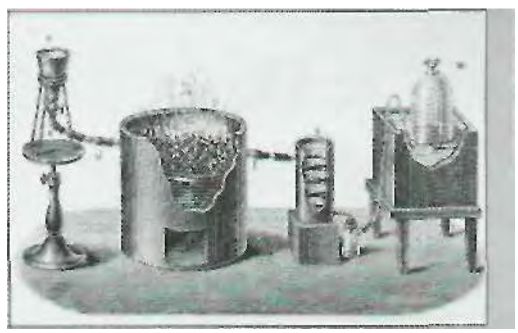
In 1802 the English chemist John Dalton (1766-1844) presented his theory of atoms to the Literary and Philosophical Society of Manchester. Although the ancient Greeks had thought of atoms, Dalton was the first to provide evidence for them, based on the relative combining weights of the elements.

Dalton's atomic theory was published in 1805. In it he said that all matter is made up of particles which are indestructible and indivisible. The atoms of an element are identical in weight and chemical properties; but atoms of one element are different from those of all other elements.

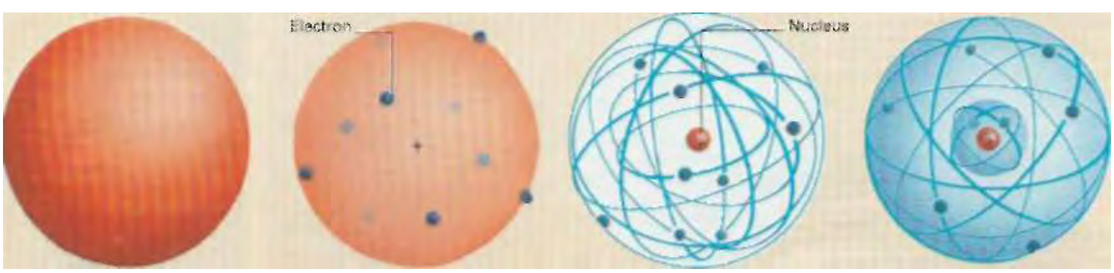
Dalton argued that when elements combine to form compounds, they do so in simple ratios, Scientists had realized that compounds had fixed compositions and that certain weights of the elements combined to form compounds; they had already drawn up a series of "combining weights". But Dalton was the first to see that elements combine in simple ratios. He also saw that when compounds decompose atoms can be released of the elements that they comprise, and that these can be used to take part in further reactions, unaffected by compound formation and decomposition.

Dalton also published atomic weights of the elements which were, unfortunately, often wildly inaccurate. But his ideas were used to calculate formulae for chemical compounds. Slowly, though, evidence accumulated that atoms were not the indestructible spheres that he had thought.

The elements of water



A *Antoine Lavoisier used this apparatus to show in 1783 that water is made up of hydrogen and oxygen. The English chemist Joseph Priestley 11733-1804) had prepared oxygen in 1774 but did not recognize its role in combustion; Lavoisier named it oxygen and showed that combustion consisted of oxidation.*



Dalton: "billiard ball" 1803 Thomson: "plum pudding" 1901 Rutherford:"electron cloud" 1911 Bohr: "shell" 1913

c- *Dalton saw atoms as solid objects like billiard balls. When Thomson discovered the electron, he imagined electrons as "plums" in a "pudding" of positively charged matter. Rutherford discovered the nucleus with electrons in orbit (t page 79); Bohr showed these orbits are arranged in distinct "shells" (t page 69).*

Until the 19th century only a handful of chemical elements had been clearly isolated

Transmutation of the elements

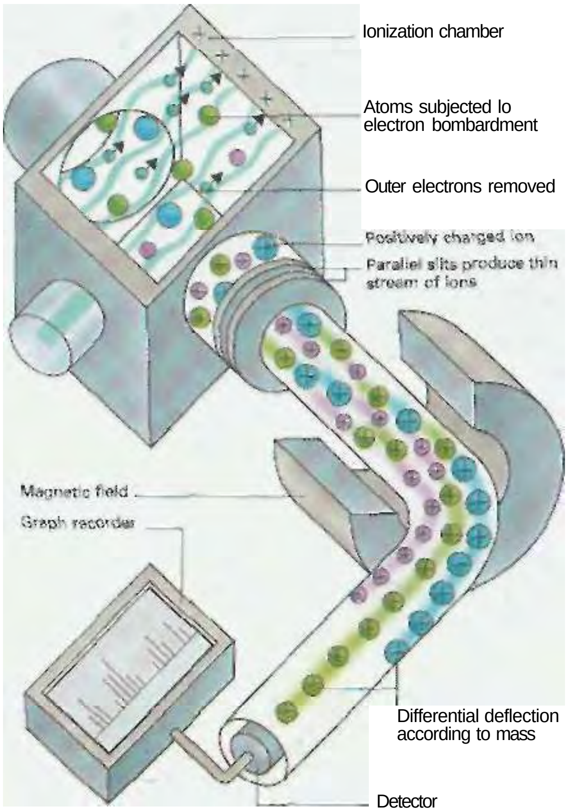
Elements can change into other elements, or transmute. Medieval and oriental alchemists were continually trying to find a way to transmute base metals, such as lead, into silver and gold. The transmutation of some metals takes place naturally, but by performing nuclear reactions in laboratory conditions the elements can also be transmuted artificially. Radioactive isotopes decay in various ways, emitting particles and energy from their nuclei. Three natural radioactive decay series are known, starting from isotopes of uranium and thorium. One series starts from the isotope of uranium which has 238 protons and neutrons in its nucleus (written as ^{238}U or U-238); the second from the uranium isotope with 235 protons and neutrons (^{23S}U); and the third from the thorium isotope ^{232}Th. All the natural decay series end with a stable, or non-radioactive, isotope of lead.

Each of these natural series occurs through the sequential decay of radioactive isotopes with the emission of alpha particles or beta-minus particles ($ page 82).

A *With their primitive apparatus, seen in this 18th-century painting, alchemists could perform only the most rudimentary chemical experiments. They searched for the elixir of life, and the "philosopher's" stone, to turn base metals into gold. Despite a few claims to the contrary, this conversion of one element to another was a hopeless task until Lord Rutherford detected nuclear decay (It page 80).*

• *Modern techniques permit the photographing of individual atoms which are only 30 billionths of a centimeter across, like this group of seven uranium atoms. The picture was taken using a scanning tunneling microscope ISTM), which was developed in the 1980s.*

The mass spectrometer



•• The mass spectrometer can count and weigh atoms and molecules. Within its vacuum chamber, a substance is ionized by a strong electric field. This knocks off some of its atoms, and some of their electrons, giving the fragments a positive charge. As these charged ions move out of the chamber and through a magnetic field they are deflected, to an extent determined by their mass and charge. A detector records the mass spectrum. Even though most molecules are fragmented, once the pieces are identified they can be reassembled to discover the original molecule.

Atomic masses

Atoms are very light, so scientists use a shorthand notation to represent their mass. This shorthand gives a scale of the masses of atoms of the elements starting with hydrogen as the lightest, but using the common isotope of carbon, ^{12}C, as a reference. Even using this notation, none of the atomic masses calculated are whole numbers. One might expect the atomic mass of carbon to be 12 on the scale, as this is the reference mass. But in fact carbon does not exist as a single isotope, and the atomic mass calculated for it represents an average of the masses of the two isotopes for carbon found naturally- ^{12}C and ^{13}C. Some 98·89 percent of carbon is found as ^{12}C and 1.11 percent as ^{13}C: if a weighted average of these percentages is taken, the figure of 12·011 is calculated. Other atomic masses are calculated in a similar manner, but relative to the mass of carbon. Not all elements occur naturally as mixtures of isotopes. The atomic masses of these elements are not weighted averages, but are simply calculated with reference to the mass of ^{12}C.

The atomic nucleus

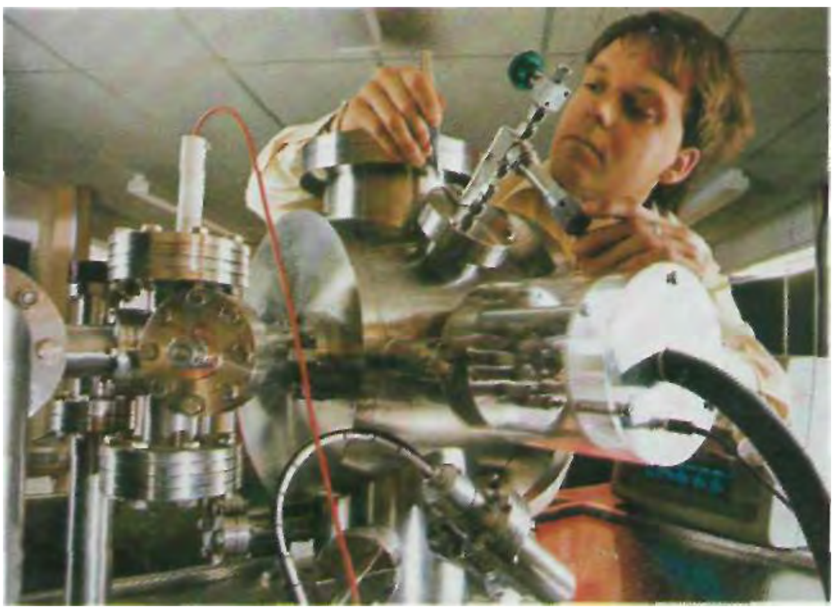
Atoms range in size from 0·1 to 0·5 nanometers (nm). Nuclei are some 100,000 times smaller, and have a diameter of roughly 10^{-6} nm. Most of the mass of the atom is concentrated in the nucleus, and is made up of the masses of protons and neutrons. The masses of protons and neutrons have been measured to be about $1{\cdot}674 \times 10^{-27}$ kilograms. An electron weighs about 9.110×10^{-31} kilograms and is 1,836 times less heavy than a proton.

The atoms of an element are identical, but different from those of all other elements. Hydrogen atoms are the simplest and lightest: they comprise a nucleus of a single proton, and one electron. The heavier elements have a more complex nucleus, with increasing numbers of protons. The number of protons in the atom characterizes it, and is known as the atomic number. Hydrogen, with one proton, has the atomic number 1, and uranium has 92. Elements with even more protons are known, but they can usually only be produced artificially by nuclear reactions.

Atoms and isotopes

The mass of an atom is another frequently quoted property. This is the sum of the number of protons and neutrons in the nucleus. The number of neutrons in the atoms of a particular element can vary, whereas that of protons cannot. Chlorine atoms, for example, always have 17 protons, but there are two distinct forms of chlorine atom known to exist naturally. In one, the nucleus contains 18 neutrons; in the other, 20.

These two forms of chlorine are examples of isotopes. Some isotopes are unstable and degrade to form isotopes of other elements, and emit radiation as they do so (^ page 83). The isotopes of some elements in particular are especially stable, and this is found when the number of neutrons in the nucleus is in the series 2, 8, 20, 28, 50, 82, 126... . The numbers in this series are known as magic numbers. A magic number of either neutrons or protons makes the nucleus of an isotope stable. In rare cases, isotopes with a magic number of both neutrons and protons are found; the lead isotope with 82 protons and 126 neutrons is one example, and is particularly stable.



The arrangement of electrons around atoms other than the simplest is too complex even for large computers to describe

Electrons and their orbitals

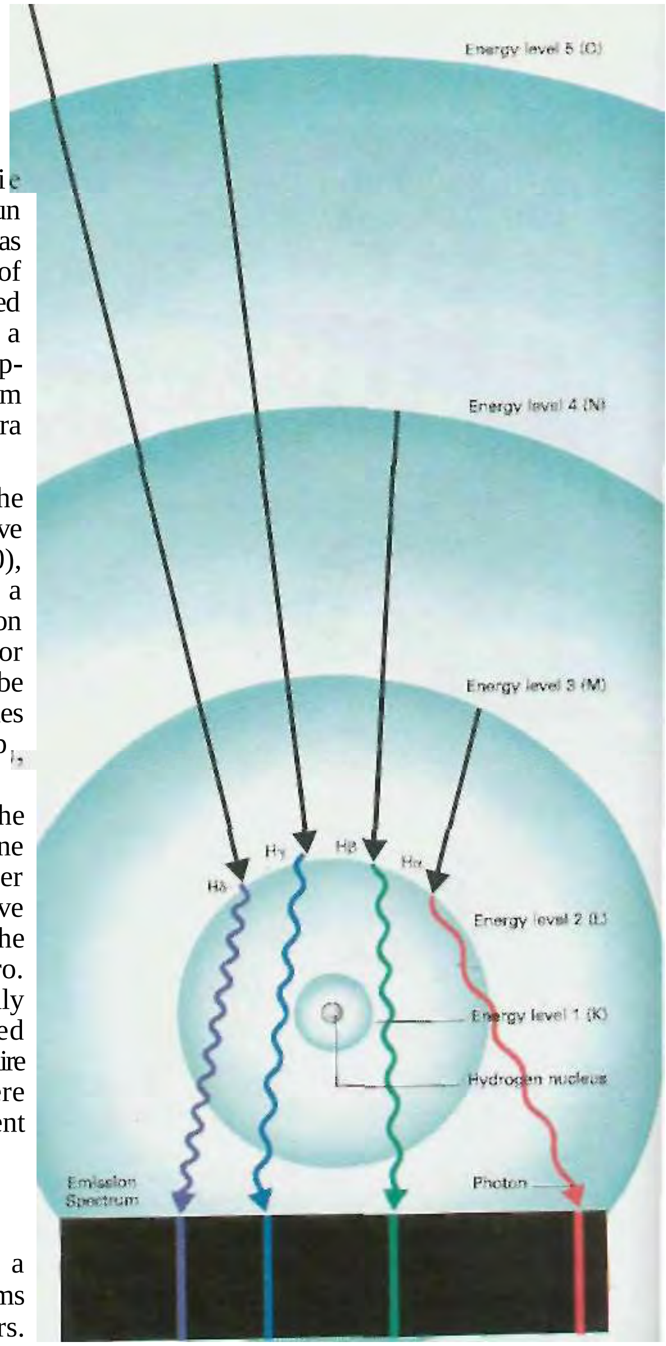
Electrons are fundamental particles with negative charge, which were once thought to orbit the nucleus much like the planets orbit the Sun in our Solar System. The Danish physicist Niels Bohr (1885-1962) was the first to suggest this, basing his ideas on the emission spectrum of hydrogen. Bohr postulated that electrons' moved in clearly defined orbits around the nucleus, and that in each of these orbits they had a certain energy (| page 88). Electrons absorb or emit energy by jumping between orbits. This theory Was able to account for the spectrum of hydrogen, but it could not be used to explain the emission spectra of the heavier elements.

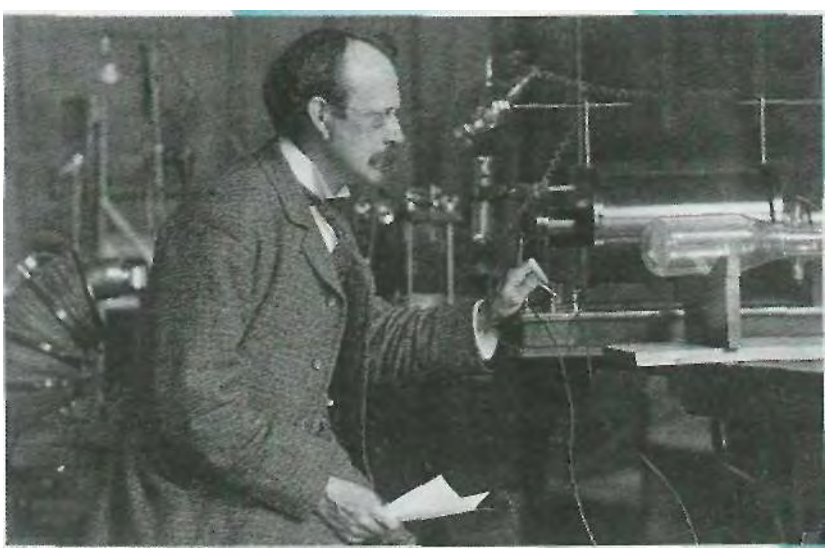
However, later work look into account his ideas and led to the current thinking on the behavior of electrons in atoms. Using a wave equation for particles devised by Erwin Schrodinger (fr- page 90), volumes of space can be mapped out within which electrons with a particular energy can be said to exist, even though their exact location cannot be stated. These volumes are known as *atomic orbitals.* For hydrogen with its single electron, a number of these volumes can be derived. Each type of volume, or orbital, is named according to lines observed in the hydrogen emission spectrum - sharp (s), principal (p), diffuse (d) and fundamental (f).

The simplest atomic orbital is spherically symmetrical around the nucleus and is known as the s orbital. Electrons with more man one energy can exist in larger spherically symmetrical orbitals or in other orbitals which have different shapes. The So-called p orbitals have twin lobes. Electrons can exist in either of these lobes, but the probability of finding an electron near to the nucleus tends to zero. There are three; p orbitals for hydrogen atoms and these are mutually perpendicular:-lying along thex, y and z cartesian coordinates. Fived orbital volumes have been calculated. These volumes again aire strongly directional, and consist of lobes around the nucleus. There are seven f orbitals with shapes which are :very hard to represent graphically.

The "buildup" principle

Aldiough the Scllrodinger wave equation for the sole electron in a hydrogen atom cab. be solved fairly easily, it is less simple for atoms with more than a few electrons, even using the largest computers.

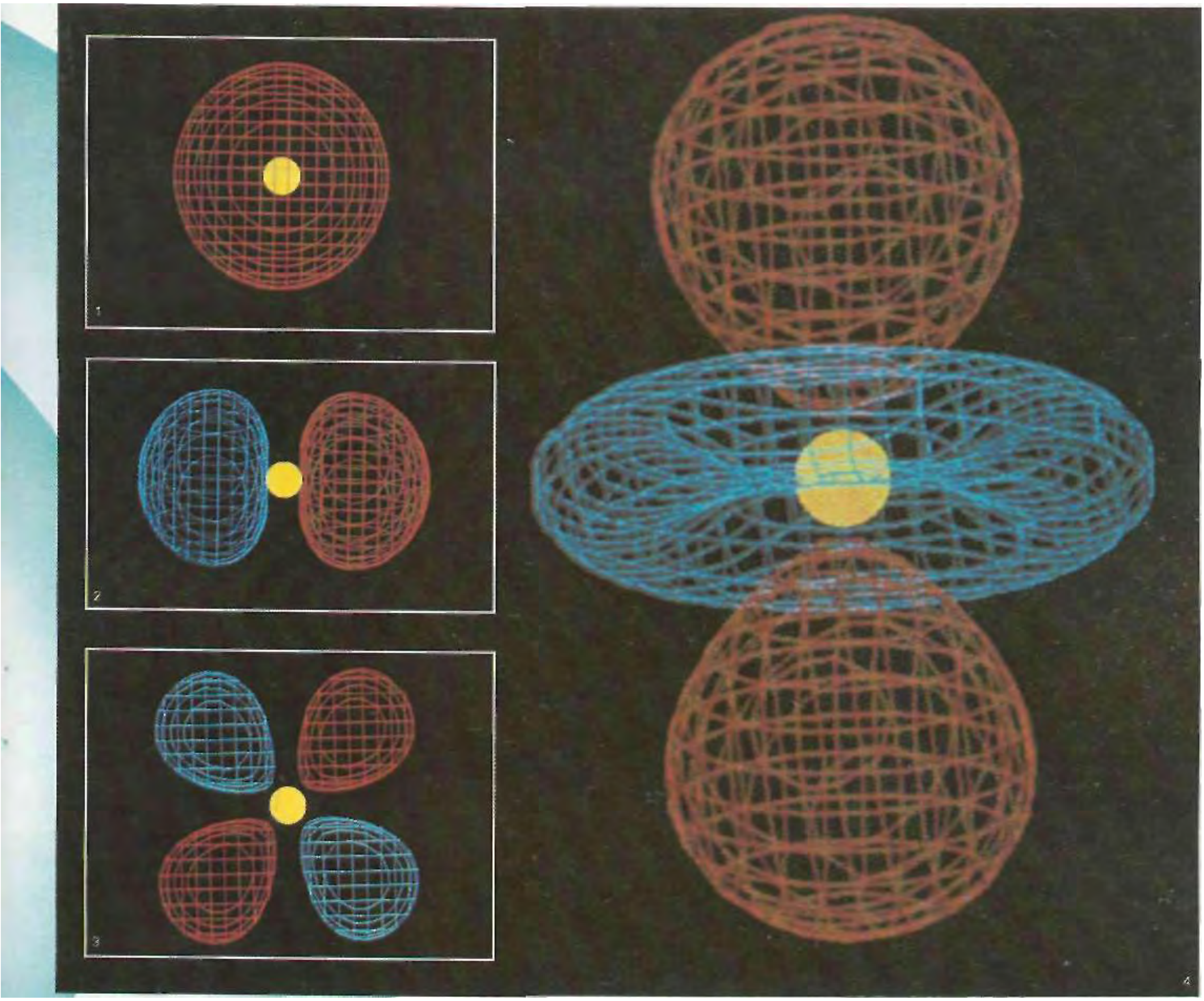




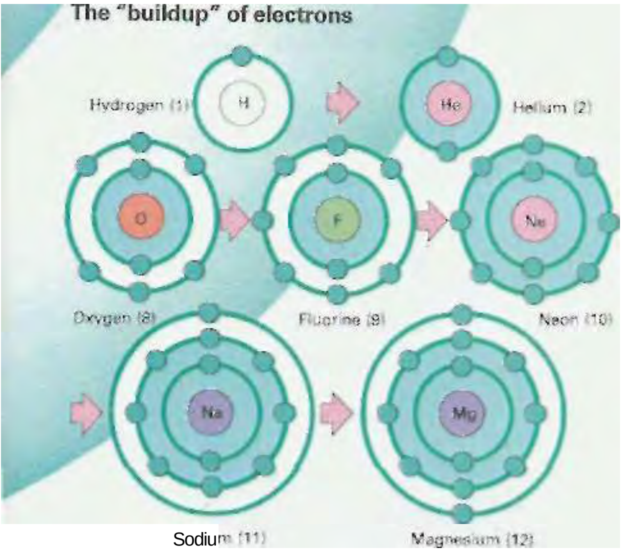
▲ *Around the hydrogen atom there are several energy levels. The sole electron of hydrogen resides in the lowest energy level IK), nearest the nucleus, but it can be excited to the outer levels. On returning to a lower level It emits a photon of light. When it returns to the L level it emits visible light where a series of four lines are found. In the ultraviolet region, lines correspond to the electron returning to its original K shell.*

◄ Joseph J. Thomson 11856-1940) discovered the electron. Using a cathode ray tube he was able to show that all elements contained these particles.

► Although electrons are particles they also behave like waves. Thus two pictures of electrons in atoms have emerged. The particle model has electrons orbiting the nucleus, like planets round a sun, and these are shown for some of the lighter elements (right). The first orbit can hold two electrons, the second eight, and so on. The wave picture gives a very different image (uppei right). The wave form of an electron depends upon a series of quantum numbers. The shapes are called orbitals of which there are four types: s spherical II); p dumbbell (2); d four-leaf clover 13) or hour-glass and ring 14); and f very complex.



The "buildup" of electrons



Even so, the general principle of atomic orbitals can be applied to atoms other than hydrogen. Orbitals similar to those of hydrogen are assumed to exist for all other elements, and models of these elements' atoms are built up by adding one electron at a time. This principle, developed by the Austrian-born physicist Wolfgang Pauli (1900-1958), is known as the Aufbau, or buildup, principle. Pauli proposed that no two electrons could exist in an atomic orbital with the same energy. However, they can exist side by side if they have opposite spin (| page 93). As electrons are added to atomic orbitals they pair up, and their spins arc mathematically canceled out. Thus each orbital can contain only two electrons. If electrons have two equal possibilities, they move into an orbital of their own. Thus the p, d and f orbitals of electrons other than hydrogen in their normal state fill individually before any electron pairing takes place. This buildup of electrons in atomic orbitals allows scientists to explain many of the chemical and physical properties of the elements, and lies at the root of the modern understanding of chemistry (| page 70).

The Periodic Table

The Periodic Table

If the elements are arranged in order of increasing atomic number, surprising correlations are found between their chemical and physical properties. Elements with similar properties are found at definite intervals of atomic number, or periods, and it is possible to draw up a table, known as the Periodic Table, which shows these similarities.

In a Periodic Table the elements are separated into horizontal periods and vertical groups. Elements in each group are found to be very like one another. Across the periods of the table, the properties of the elements are seen to change steadily. The table shows the distinction between metallic and nonmetallic elements and, when first introduced, it allowed chemists to predict the existence of elements which were then unknown. The Periodic Table includes all the naturally occurring elements but, by extrapolating from the known elements, it can be used to predict the properties of as yet unknown superheavy elements.

The origins of the Table

The Periodic Table originated from the desire of chemists in the 19th century to present the observed similarity in chemical and physical properties of the elements in a systematic way. By the middle of that century almost 70 of the 92 naturally occurring elements were known. Substances such as iron (Fe), tin (Sn), lead (Pb) and gold (Au) had been known as elements for some time, but it was very hard to prove thai an element was actually an element. It was easier to show that an element was not an element than that it was. Only with the advent of spectroscopy and modern laboratory analytical techniques has conclusive proof been available.

However, the Russian chemist Dimitri Mendeleev (1834-1907) first illustrated the periodic nature of the elements by placing them in order of increasing atomic weight. In his periodic law, he said that the properties of the elements vary in a systematic way, and are related to the atomic weights. Mendeleev arranged all the elements he knew into a table of eight vertical groups and 12 horizontal periods. Often he left gaps when the atomic weight of an element did not fit his scheme. These were later filled in by newly discovered elements. For example, he predicted that an element which he called ekasilicon should exist, placed in the table below silicon (Si) and above tin (Sn). This was later discovered, and is now known as germanium (Ge); it occupies the place in the table that Mendeleev predicted, and has physical properties similar to both silicon and tin. Mendeleev's work was very accurate, and led to the search for new elements and for improved values of atomic weight and density of known elements. Several new elements have been discovered, and predictions made of new transuranium elements.

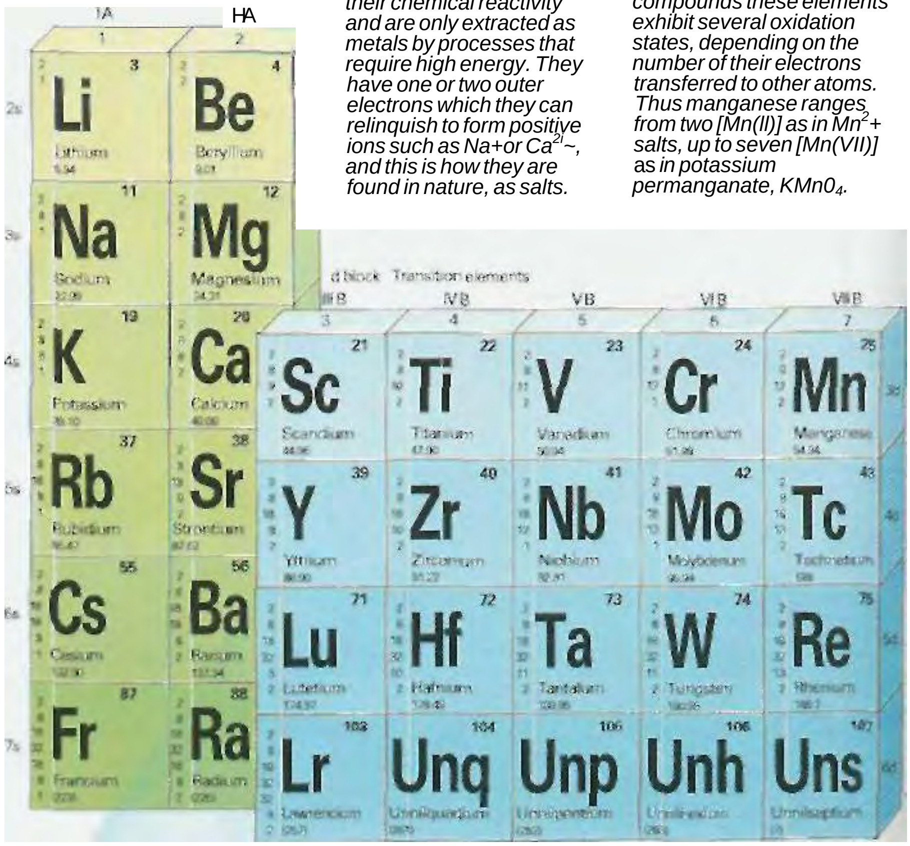
Nevertheless, his ideas were not entirely correct, as an element's properties depend on its atomic number- the number of protons in the nucleus - rather than the atomic weight. The periodic nature of the configuration of the outermost electrons of atoms explains observed similarities in chemical and physical properties.

•« The s block consists of group 1, the alkali metals, and group 2, the alkaline earth metals. These elements are known for their chemical reactivity and are only extracted as metals by processes that require high energy. They have one or two outer electrons which they can relinquish to form positive ions such as Na+or $Ca^{2!}$~, and this is how they are found in nature, as salts.

• The d block elements are all metals, in which the d shell, underlying an outers shell, is being filled with electrons. In their compounds these elements exhibit several oxidation states, depending on the number of their electrons transferred to other atoms. Thus manganese ranges from two [Mn(II)] as in Mn^{2}+ salts, up to seven [Mn(VII)] as in potassium permanganate, $KMnO_4$.

s block

HA



f block Inner transition elements

57 58 60 61 62 i

Ce Pr *Md* Pm iSm-Eu

! Lanthunui | t) Cerium 2 iau.12 | 2 Praseodymium 1403 | I 2 Neodymium ! 144. ?4 | 2 Promelhiufn | 2 Samarium | 2 Europiunj

90 92

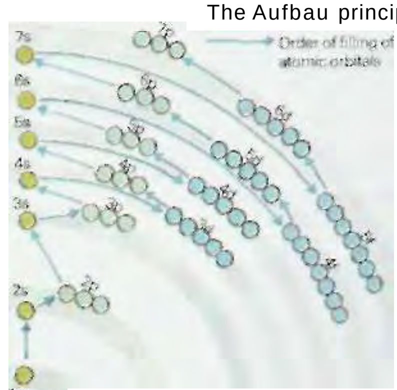
Ac T!h Pa JU Np iPu :Am

9 Actinium ?. ton | 9 Thorium 2 23204 | 9 Protactinium | o Uranium | 9: Neptunium U28?L | .23 Plutonium •LAJ2SB | W8 Amoricjum- / I7OT

•• The Periodic Table reflects the electron make-up of the atom. Electrons occupy shells around the nucleus. The inner shell can hold two electrons, the next eight, and so on, up to 32 for the outer shells. Within these shells are subshells of 2,6, 10 or 14 electrons with the symbols s, p, d and f- terms originally from lines in the atomic spectrum. The Aufbau principle sees the table built up by a step-by-step addition of electrons to atoms, from the nucleus outwards. When a shell is full it results in an atom of great chemical inertness - the rare gases.

Energy level

The Aufbau principle



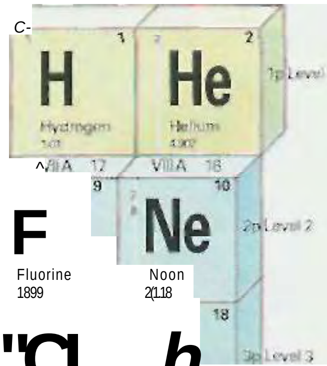
1s

Maximum number ot electrons tn each shell

2 8 8 18 18 32 32

•< Dimitri Ivanovich Mendeleev was born in Tobolsk, Siberia, in 1834. His success in winning a scholarship to study in Germany put him among leading chemical thinkers of the day. He returned to a teaching post at the University of St Petersburg (now Leningrad) where he discovered the Periodic Table in 1869, while writing a textbook of inorganic chemistry. He died in 1907.

ishlock C-



p block

IV A

15 / 16 / ^/IIA

N 0 F

Nitrogen 1401 Oxygon H.99 Fluorine 1899 Noon 2(1.18

P $ "Cl h

Phosphorus 3097 Sullur 32.06. Chlorine 3516 Argon 39.06

VIIIB

/ r ^ 7

Fe Co iNi Cu Zn tfS ^§e

Cobalt K93 Copper 6355 Zinc fcfi.37

As &e Br :Kr 4piLevo 4

Atsenic as Selenium «tfc Bromine WW Krypton .83.80

4S 48

51 52 j 53 j ,

Ru iRh iPd ;Ag =Cd 4d' 3 In Sn

Indium IMS? Tin

Palladium Cadmium

Sb Te =| iXe ; 5p* Level 5

Antimony 171.75 Tellurium 127.60. • iodine 12630 Xenon totJC

TO j 3 7? 2 78 .. 79 i , 80

83 j 84 .. 85 -B6j^

:0s !:'lr sPt iAu -Hg |5d is 3 Tl Pb

Bi -Po ;At ;Rn 6p;Leve 6

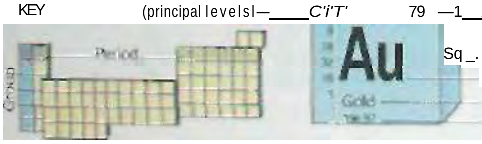
in riejiiirn., Platinum Gold 2 Mercury Thallium 20437. Lead »7iia

1 Bismuth .20898 6 Polonium 17101 1 Astatine •{?»» 8 Radon (222)

KEY

Electronic configuration (principal levels) 11 — Umip ruimhrr

79 —1 Atomic number



Sq . Orbital designation
Chemical symbol
Name of element
Relative atomic mass

j, ,J ,,r ~^

64 I i 65 I 2 6 2 7 ? 6 ! ? 69 I 2

:Gd iTb IDy 3H0 iEr iTm jYb

Gadolinium Terbium * Dysprosium * Holmium Erbium Thulium * Ytterbium

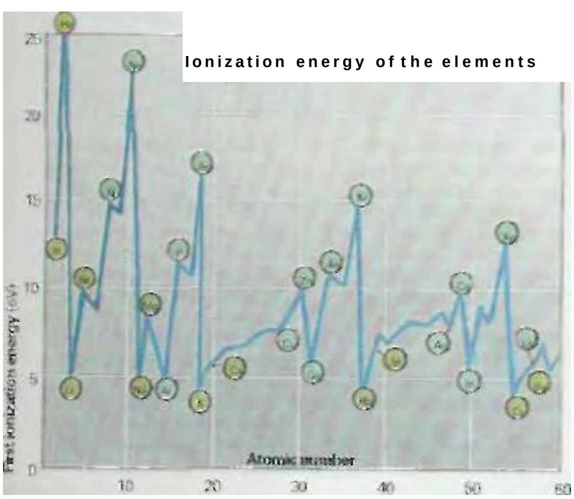
96. 9? • 98 99 160 tor loz j

CrrUBk |Cf 3Es iFm iMdpto

t Cu'iiinr 28 ri Bckelium 1.2?! 9 Californium 28. 9 Einsteinium 29 9; Feunium 30 u Mendelevtum 3V a Nobalium

•« The f block metals consist of the lanthanides (top row), and the actinides (bottom row}. The lanthanides are also called the rare earths, although some are quite abundant (cerium is more common than lead). They are chemically very similar, a fact that led to much confusion among chemists of the last century. The actinides are all radioactive elements, but thorium and uranium have sufficiently long half-lives that deposits of these elements are still found. The elements beyond uranium are all artificial.

A The p block is divided diagonally into metals and nonmetals. The nonmetals bond to one another by electron-sharing to form discrete molecules- the basic components of matter. Their chemical behavior is determined by the completeness of their shells. The elements of group 18 have a filled shell and are most stable, while those of group 17, the halogens, are extremely reactive. The 1s block, hydrogen and helium, are at this end of the table because both are nonmetals and helium is an inert gas.



•4 Elements arranged in a line according to their atomic number show repetition of certain properties. This is illustrated by their ionization energies - the energy needed to pluck the outermost electron away from an atom. Some elements are very loath to give up their electrons, especially if they have a full shell. Thus in the graph the pinnacles are all capped by the rare gases. The s block elements of group 1 offer the least resistance to losing their sole outer electrons, and so these are at the low points.

Using the Table to predict new elements

The Periodic Table has always been useful in predicting the existence of as yet unknown elements. Fourteen artificial, or transuranium, elements have been discovered. Elements with the atomic numbers 93-103 are the remaining elements in the actinide series; elements 104-118 begin a new transactinide series: in an expanded form of the Table these may be placed under the row beginning with hafnium and ending with radon.

The chemical properties of these transactinide elements should be similar to those in the groups of the representative elements in which they have or might be found. Around element 121 it has been predicted that a "superactinide" series might start. In this series, which would be similar to the actinides, the g orbital would fill with electrons.

See also
Molecules and Mailer 25-34
Studying the Nucleus 79-86
The Quantum World S7-96

The structure of the Table

The modern Periodic Table represents three distinct types of elements, called representative elements, transition elements and inner transition elements. The representative elements are perhaps the most familiar. They include both metals, such as tin (Sn) and lead (Pbj, and nonmetals such as oxygen (O), sulfur (S) and chlorine (CI). The transition metals are a series of metallic elements, including iron (Fe), nickel (Ni), copper (Cu), zinc (Zn) and gold (Au), with properties that are very similar to one another, but which definitely change as the atomic number rises. The inner transition metals, sometimes called rare earths or lanthanides and actinides, have very similar properties and have often been difficult to separate and identify.

The groups of the Periodic Table may be numbered 1 to 18, or the representative elements arranged into eight vertical groups numbered I to VII and O. The transition metal groups can also be numbered I to VIII. The elements are arranged into seven horizontal periods.

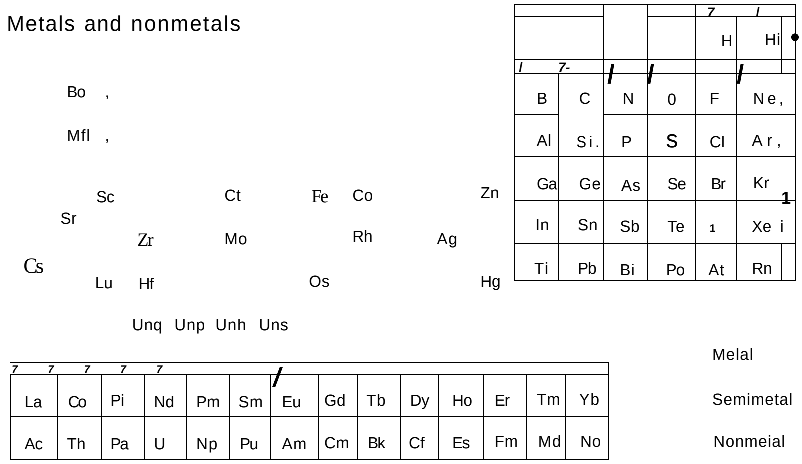
The group 18 elements - helium (He), neon (Ne), argon (Ar), krypton (Kr), xenon (Xej and radon (Rn) - lie to the right of the table. Collectively they have been called the noble or inert gases, because of their lack of reactivity; now they are more often called the rare gases. These elements are unreactive because their atoms have a complete outer shell of electrons, filled with eight electrons. This is an extremely stable structure and it is very hard to remove electrons to allow chemical bonding and the formation of compounds.

To the left of the table lies a group known as the alkali metals - lithium (Li), sodium (Na), potassium (K), rubidium (Rb), cesium (Cs) and francium (Fr). All of these are very reactive. The atoms of each contain one s-orbital electron in the outermost electron shell; this electron is easily lost, giving rise to a positively charged atomic particle or ion with the structure of an inert gas. This can take part in chemical bond formation. Group 17 comprises fluorine (F), chlorine (CI), bromine (Br), iodine (I) and astatine (At). Their atoms have almost full outer p orbitals containing five electrons. They can easily gain an electron to form a negatively charged ion, again with the structure of an inert gas. Such ions can also readily take part in bond formation.

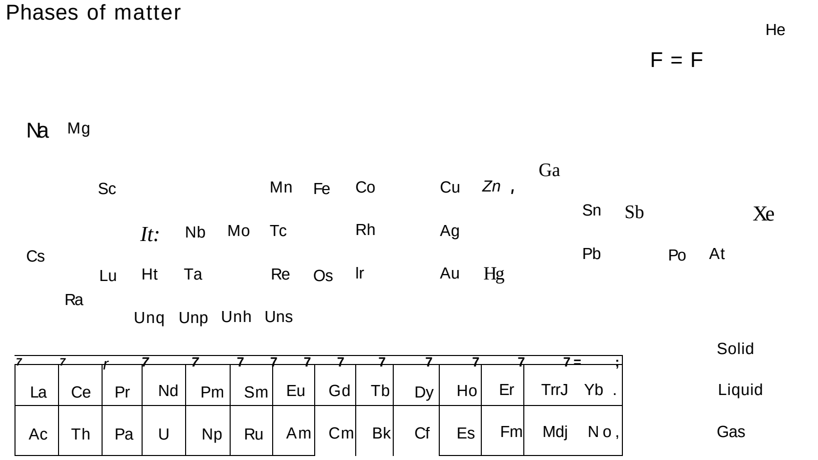
This gain and loss of electrons to form ions is shown across the periods of the table. Elements are grouped together because of their physical and chemical similarities, and their ability to form ions with the same positive or negative charge. The ability of an atom to lose or gain an electron, its "ionization energy", increases along the periods' and decreases down the groups, but it is only one of many properties that change "periodically".

Most elements are metals, but groups 13 to 16 contain elements neither metallic nor nonmetallic in character-the semimetals. In each of these groups the metallic nature of the elements increases down the group. In group 13, for example, boron (B) is a semimetal whereas aluminum (At), gallium (Ga) and other elements in the group are all metals. The other semimetallic elements are silicon (Si) and germanium (Ge), arsenic (As) and antimony (Sb), and tellurium (Te).

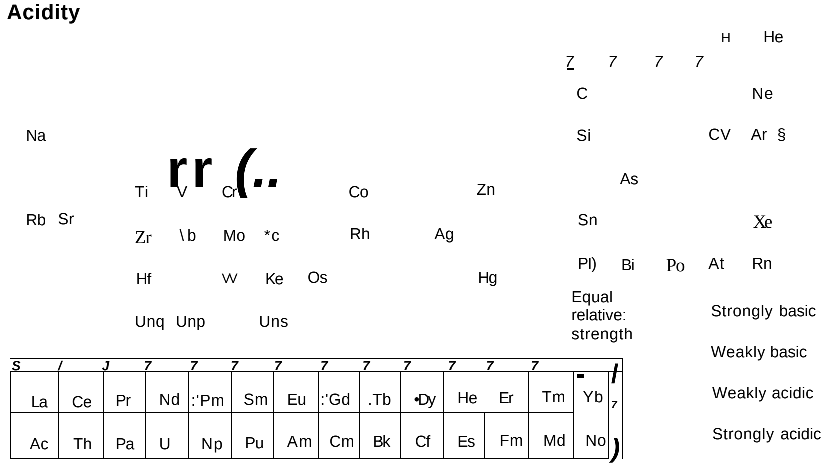
Metals and nonmetals



Phases of matter



Acidity



▲▲▲ *Recognizing a metal is fairly easy. It is hard, dense, shiny and gives out a characteristic "ping" when struck. Chemists, however, define a metal by its ability to conduct electricity. Conductivity stems from the free moving electrons that are a feature of the chemical bonding in metals. A few elements are semi-conductors, and these straddle the boundary between the metals and non-metals.*

▲▲ *The three states of matter, solid, liquid and gas, are very dependent upon temperature and pressure. At 25° Celsius and atmospheric pressure, the conditions that humans find most agreeable, only two elements in the middle table are depicted as liquids - bromine and mercury. At 30° Celsius cesium and gallium would also have been shown as liquids. These melt at 28-5° and 29-8° Celsius respectively.*

▲ *Almost all the elements form oxides, and how these behave in water is a guide to the chemistry of the element. If the oxide reacts with water to give an alkaline solution then it is described as a basic oxide; if it reacts to form an acid solution it is described as an acidic oxide. Metals give basic oxides, non-metals give acidic ones. Amphoteric oxides, such as aluminum oxide, show both kinds of behavior.*

Using the Elements

10

Iron and steeL.Other alloys...Making liquid oxygen-Separating the gases in the air...Silicon and germanium, the semiconducting metals...PERSPECTIVE... Metallography and archeology...Using helium, neon and argon

Materials found in nature have always been used to make life easier and safer. Over some 7,000 years, wood, bone and flint were the raw materials used for most artefacts. Since that time the elements from nature have been separated and recombined and fashioned into an almost endless list of everyday objects, tools and machines.

The separation and isolation of some of the elements on a large scale has driven the advance of technology since the Industrial Revolution. At the forefront of this progress has been the use of metals and alloys to build better and more efficient machines.

Iron is the most widely used metal. It is produced in three main types - wrought iron, cast iron and steel. Pure iron (Fe) is a white lustrous metal that is quite reactive and not very hard. It is attacked easily by air; iron hydroxide is formed and flakes away exposing more metal. But when iron is made on a large scale in a blast furnace, it mixes with carbon and silica (S1O2) to give irons of varying degrees of hardness and malleability. Iron is drawn from a blast furnace in ingots or pigs and these can be used to make the three main types of iron - wrought iron, cast iron and steel.

Wrought iron is essentially a malleable alloy, or mixture, of iron and phosphorus with small percentages of carbon, manganese, sulfur and silicon as impurities. The properties of wrought iron depend upon all these ingredients and also upon the amount of glassy fibers of slag produced from the "scum" or waste material - including silica - in the blast furnace. Cast iron contains between 1-8 and 4-5 percent carbon plus elements such as silicon, phosphorus and manganese.

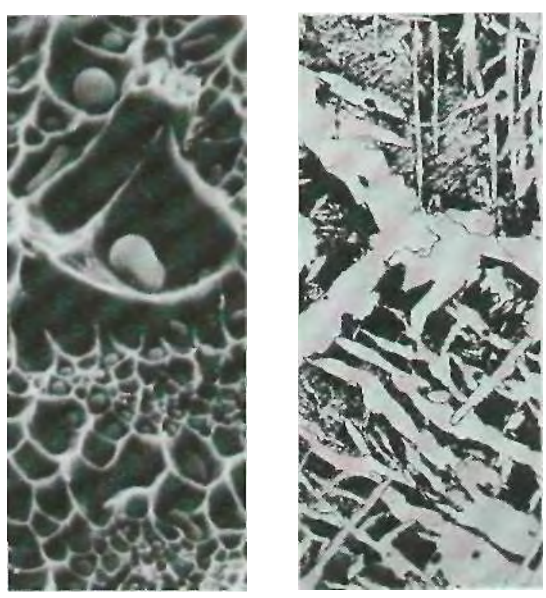
The properties of cast and wrought types of iron can be changed by altering the conditions under which they are made and by reducing, or increasing, the amounts of impurities. Commercially pure iron can be prepared in either a specially-lined blast furnace, or by electrolysis (4 page 56), but it is a very different material from pure iron. This type of iron contains only about 0-002 to 0-008 percent silicon, sulfur, phosphorus, carbon and manganese.

Iron alloys are usually produced from cast iron. Different percentages of different elements give these alloys better mechanical and physical properties when in use. The most widely used alloy of iron is steel. Steel is really an iron alloy with a low carbon content, the alloying elements being manganese, cobalt, chromium, molybdenum, nickel, tungsten, vanadium, selenium, aluminum and other metals in varying combinations for different purposes. Chromium and nickel are added to steel to improve its corrosion resistance and malleability. Very hard tungsten steels are used in high-speed cutting tools.

Stainless steels contain, relatively high percentages of nickel and chromium, but other elements are often added - including bismuth, selenium, silver and lead. The addition of these elements helps make the machining of stainless steel parts easier.

• H W H I

The Pillar of Mehauli (near Delhi) is a column of pure iron over 10m tall. It has resisted rusting for nearly 1500 years, but this is less due to any special metallurgical knowledge of the ancient craftsmen, just a favorable climate. The heat of the day is retained by the pillar sufficiently to prevent dew condensing on it during the night.



Electron microscope photographs of metal surfaces. On the left a piece offerrite (iron) that has fractured as a result of being overstretched. The round particles are actually manganese sulfide and each is about a millionth of a meter in diameter. On the right is a piece of steel magnified a thousand times.

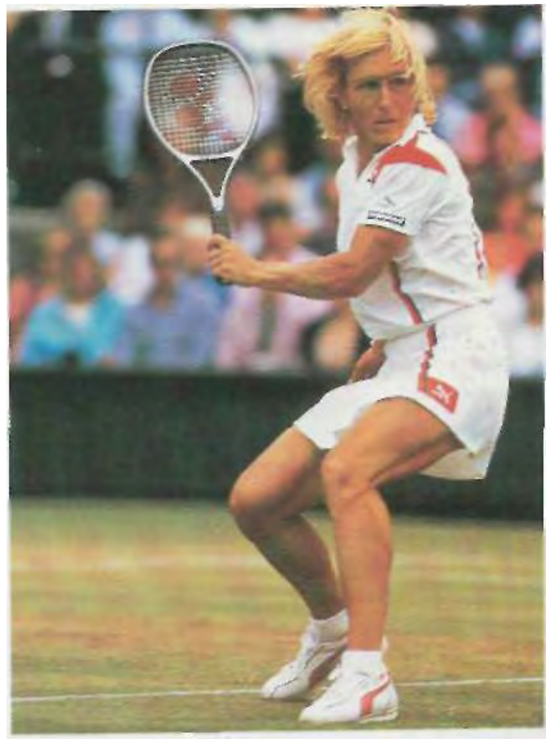
Metals, alloys and carbon

Other metals are used throughout industry, either in their pure state or as alloys. Copper can be made in commercial quantities to a purity of 99-95 percent for use in power cables, busbars and windings for electric motors and dynamos. The brasses are alloys of copper and zinc with manganese, aluminum, nickel, phosphorus and antimony sometimes added. Alloys of copper and tin, often with other elements added, are called bronzes.

Special applications require special materials and, while steel still provides the bulk of all engineering materials, other metals and alloys are used to make aircraft, spacecraft and some automobiles. Titanium is used to make aircraft but, as with steel, engineers usually find that titanium alloys are more useful than die pure metal. Aluminum is also used whenever corrosion-resistant structures are needed.

Carbon is used in industry as a pigment, additive and in the manufacture of inert coatings and electrodes. Small fibers of a polymeric form of carbon can be made, and these are very strong. These carbon fibers are used to make lightweight but strong composite materials for use in aircraft structures and many leisure goods such as tennis rackets and skis. Because carbon is inert the fibers are also being used in medical research as implants in the body; to replace tendons for example, and in parts of heart pacemakers.

Much of the carbon produced in industry is made as carbon black. This type of carbon is made today from petroleum and tar residues, in a process that can be traced back to antiquity. To make carbon black, petroleum and tar residue "fuels" are burned in an atmosphere with insufficient oxygen. A sooty deposit of carbon is formed, which settles and is collected. The types and sizes of particles formed from the carbon black process play an important part in the uses to which carbon black can be put. Most carbon black is used in the rubber industry as a reinforcement, particularly in the manufacture of automobile tires. A loose or "fluffy" form is used as a pigment, mostly for newspaper printing inks but also for paints. Some special forms of carbon black are added to plastics to make them electrically conductive or to improve their ability to resist weathering and their antistatic properties. Because of its high electrical conductivity, carbon black is also used to make brushes for dynamos and electric motors and also to make electrodes.



•4 Titanium is a very light metal; its alloys are strong and have a high melting point which makes them, suitable for use in supersonic aircraft, where friction from air resistance is high. Recent experiments have been made to test its suitability for prostheses (artificial limbs); its strength, lightness and resistance to corrosion are its main advantages here.

.•: Carbon fiber has recently found many new uses, particularly for objects that need to combine lightness with strength. Many leading tennis players prefer carbon-fiber rackets (top), while carbon-fiber reinforced plastics have revolutionized light aircraft design (above); as much as 60 percent of such an aircraft by weight may be made of this material.

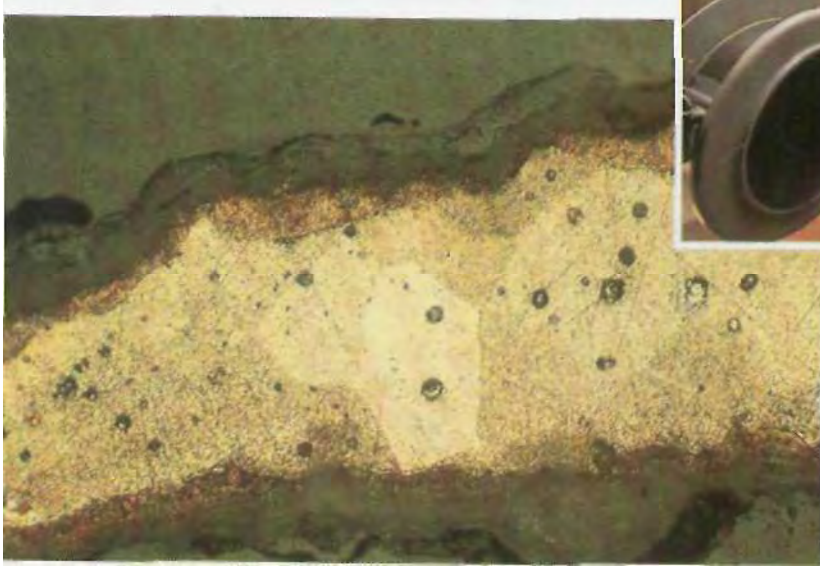
Chemistry and archeology

The skills of the chemist, particularly those of the metallurgist, can come to the aid of the archeologist in unraveling the achievements of early cultures.

The figure shown left, about 20cm high, was produced by the Indians of Colombia. The main material is a gold-copper alloy, which was made simply by melting gold and copper together to combine hardness with brilliance. Study of the micmstructure of the metal in an artefact can reveal much information about the way in which it was made. Analysis of minute cross-sections of the metal has shown that the surface layer was treated with acidic plant juices to remove an ultrathin layer of copper. This technique, known as depletion gilding, left a bright sheen on the metal.

The object was made to be hung around the neck, and contained lime. This was used by the Indians to release the alkaloids in the coca leaves that they chewed in order to stave off hunger and induce altered states of consciousness.

In addition to metallographic analysis, spectrography. X-ray fluorescence analysis and examination of samples of the metal under the optical and electron microscopes, it has also proved possible to subject this particular piece to radiocarbon dating. The statuette was made by the lost-wax technique whereby a wax model was surrounded with clay and then baked. The wax was run off, leaving a clay cast into which the molten metal was poured. Tiny amounts of clay remain in the core, and these have been analyzed and the piece dated to between AD 600 and 1000, 500 years before the arrival of the Europeans.



-< A The gold figure was collected in Colombia, South America, and has been subjected to metallographic analysis to confirm its date and method of manufacture. The micrograph of a section through the metal shows the very thin surface layer, where the copper in the gold-copper alloy has been carefully removed to enhance its appearance.

A An archeologist submits another ancient artefact, this time a Roman coin, to metallographic analysis. By placing the coin in a mass spectrograph, the metal content of the alio yean be determined. Knowledge of the proportion of precious metal in coins is invaluable information to the historian attempting to draw a picture of the economic conditions of the past.

Neon, helium and argon are useful precisely because they do not react easily with other elements

Making pure oxygen

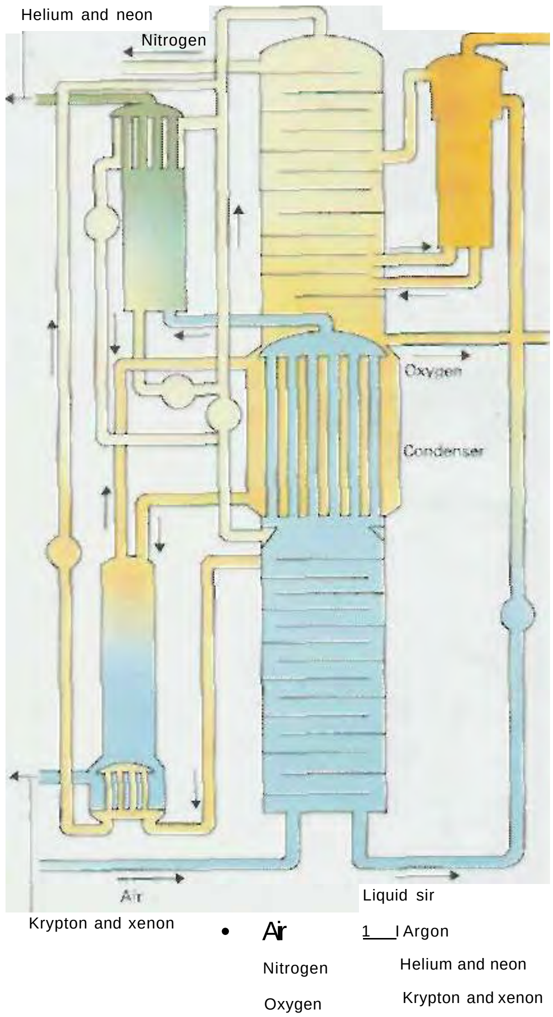
Oxygen is used in many of the process industries to make combustion more efficient. It is vital to the cost effectiveness of many chemical and metallurgical processes. Oxygen is used in the production of both iron and steel; in the chemical industry it is used to "crack" methane (natural gas; CH_4) to produce acetylene (C9H2) and also in the manufacture of oxygen-containing compounds.

Air contains almost 20 percent by volume of oxygen and is the raw material for almost all oxygen production. Air can be separated into its component gases if it is first liquefied and then distilled. The various gases in' the air - oxygen, nitrogen and some argon - can be separated in this way because of their different boiling points. Liquid oxygen is used as a propcllant in missiles and rockets and also to produce oxygen for breathing in aircraft, spacecraft and hospitals.

Two basic processes are used to liquefy gases and widely used in all oxygen-producing plants. In both processes air is first filtered to remove dust and other solid particles and then compressed. This compressed gas is further purified if necessary and then cooled, both with the help of refrigerant chemicals and by the exchange of heat with recycled gas, product and waste gas from the plant. In one process the compressed gas is allowed to expand through a throttle valve or nozzle. The volume of the gas before and after expansion is kept constant, heat is lost and the gas cools. Some of the gas is cooled sufficiently for it to liquefy, and this is either stored or passed to a distillation column to be separated into oxygen, nitrogen and argon.

In the second method, the liquefaction process is more efficient. Cooling, is achieved by expanding the compressed air through an engine, usually a turbine, doing what is called "useful work". Expansion of gas through the engine is at constant entropy (4 page 34), energy is removed from the system and the temperature of the gas drops. The engine does not produce liquefied air, but expansion through a valve and further cooling allows liquid air to be made.

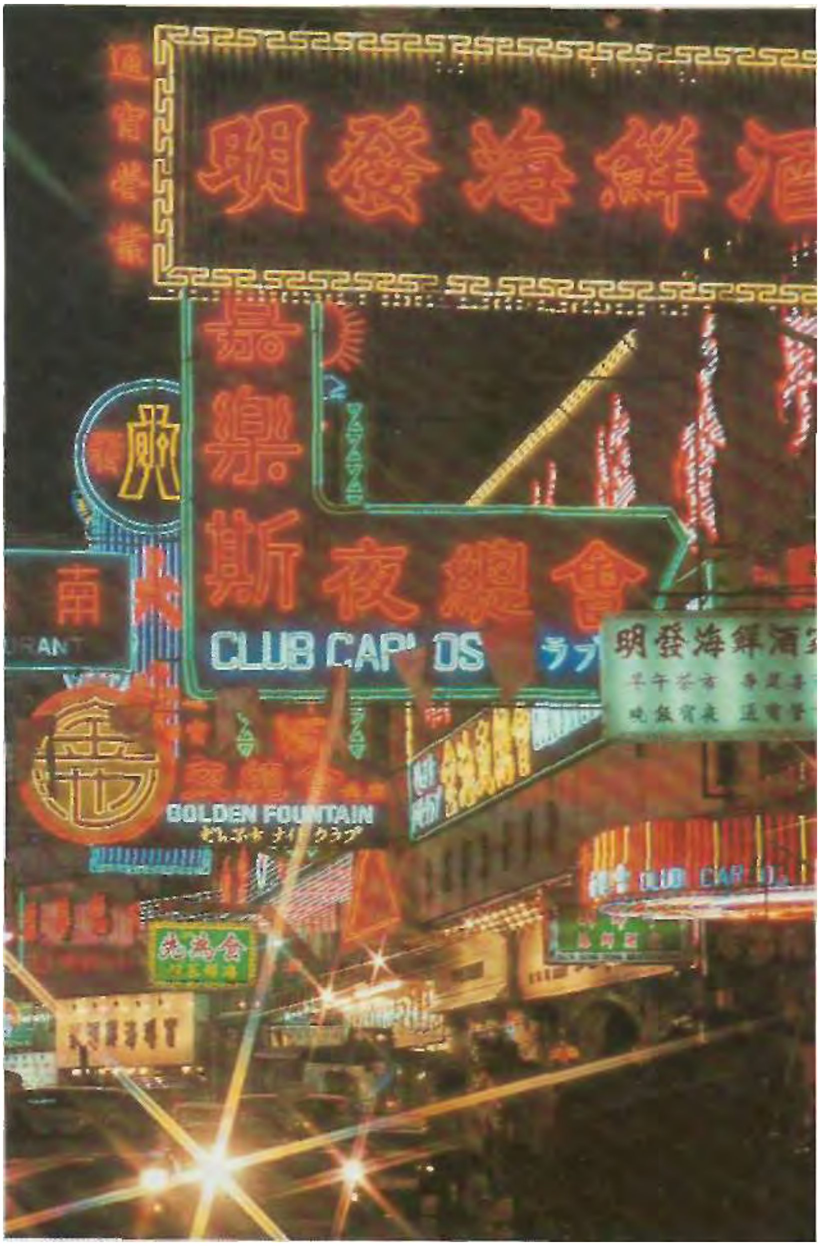
When liquid air is warmed the first gas to be given off contains 93 percent nitrogen and 7 percent oxygen to leave a liquid rich in oxygen. As the temperature is raised towards the boiling point of oxygen (—183°C), more oxygen is released. The last liquid to boil contains about 45 percent oxygen. Because of this, oxygen separation plants use two fractional distillation columns, one at a relatively high pressure, to ensure the final liquid is almost pure oxygen.



• Six gases are extracted from the air by cooling it until it liquefies and then distilling it. Each of the component gases boils at a different temperature. Nitrogen, oxygen and the noble gas argon are the main industrial products, and are produced in very large quantities, but the rarer gases that are extracted, neon, krypton and xenon, also find commercial and scientific use.

•< Provided the water leaving a sewage works contains enough dissolved oxygen it can safely be returned to the environment at the end of the treatment. The purification of water is one of the major uses of purified oxygen.

• When the gas acetylene is burned in pure oxygen, enough heat is given out to melt and weld steel even in the coldest climates, like here on the Alaska pipeline. Sometimes the inert gas argon is used to protect metals such as titanium and zirconium from being attacked by the air while they are being welded.



Helium, argon and neon

Not all the so-called "rare" gases are particularly rare (4 page 70). Helium is produced on a quite large scale from some natural gases. Argon is separated from air in much the same way as oxygen and nitrogen. The other noble gases, krypton, xenon and neon, ate also obtained from air by liquefaction but special techniques are used to separate them.

These gases are used wherever a chemical, metallurgical or experimental process needs to be carried out in an inert atmosphere. Argon is most widely used for this purpose. It is used along with helium to provide an inert atmosphere under which some metals that are readily attacked by air can be welded. These "hard-to-weld" metals include aluminum, bronze and copper and some stainless steels; but the technique known as gas shielded arc welding, is most widely used for welding parts containing titanium and zirconium. The ability to weld these metals has increased their use in aircraft, spacecraft and in the nuclear industry.

The most common use of the inert gases is in the manufacture of "neon" lights. Neon emits the familiar orange light while krypton emits a pale violet light. Mixtures of the inert gases are used in incandescent and high-output lamps to give a range of colors and to improve their useful life.

Helium is used today on a relatively large scale to fill balloons and dirigibles. It is also used in rockets to replace liquid oxygen as it is burned and in the gas mixture breathed by deep sea divers. Helium boils at -268-93°C, or4-2TK, a temperature very close to absolute zero. Because of this, liquid helium is used to investigate the properties of matter close to absolute zero ($ page 94).

• Hong Kong glows with the colors familiar to every modern city. The inert gas neon is the secret behind the illuminated signs that brighten cities at night. When an electric discharge is passed through neon it glows red. Other colors may be produced by traces of mercury vapor.

• Helium, being an inert gas and much lighter than air, is the ideal gas for use in balloons, whether for meteorological purposes or to assist in the study of the upper atmosphere or of cosmic rays. Such balloons may be required to reach heights of as much as 30,000m above sea level.

See also
Electricity49-56
Atoms and Elements 65-72
Nuclear Fission and Fusion 119-24

The semi metals

Some of the elements in Groups 13 to 18 of the Periodic Table (fpage 71) behave neither as metals nor as nonmetais. They are not true conductors of electricity as are the metals, nor true insulators like the nonmetais. Because of this, they are useful in the fabrication of electronic devices.

The dramatic growth of the electronics and microelectronics industry followed the invention of the transistor in 1947 and more recently the silicon chip. These developments were dependent upon understanding and increasing production of semiconducting elements (ipage 54). Silicon is the most widely used such element today. Single crystals of silicon are "pulled" from crucibles containing the molten element to produce wafers for the manufacture of silicon chips.

To create a silicon crystal of sufficient purity, metallurgical grade silicon, 98 percent pure, is used. This is reacted with hydrogen chloride to form a liquid, trichlorosilane, which is distilled several times to reduce impurities to a level of parts per trillion. The trichlorosilane is then heated and decomposes to form perfectly pure silicon crystals up to 125mm across.

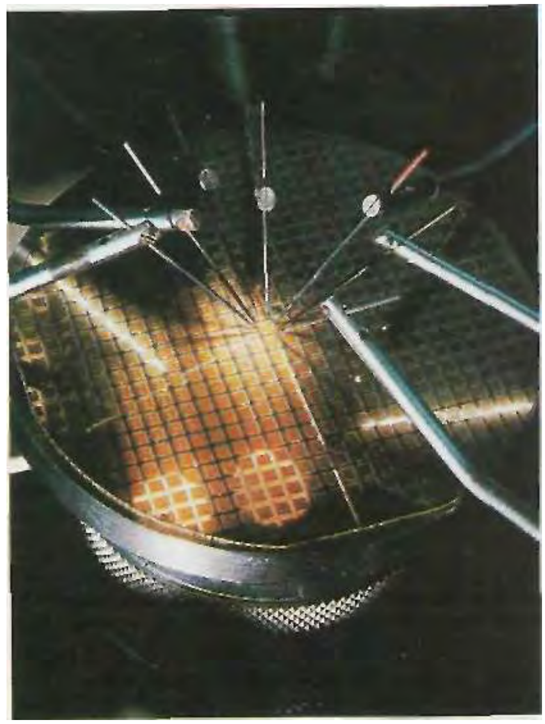
Making integrated circuits

The silicon crystals are "doped" to produce semiconducting material with special characteristics. Doping silicon with tiny amounts of phosphorus or arsenic produces "n-type " material while doping with boron produces "p-type" material. Combinations ofn- and p-type semiconductors produce so-called p-n junctions. These junctions are used to create diodes and transistors in microelectronic circuitry. Photographic and chemical techniques are used to etch wafers of silicon and other semiconducting materials to form these junctions and make ever more compact electronic devices.

Compounds or alloys of two or more semiconducting elements have unique electrical properties. One of the most exciting of these is gallium arsenide (GaAs). Because it is made from gallium, in Group 13 of the Periodic Table, and arsenic, in Group 15, gallium arsenide can be grown on a base, or substrate, of one crystal lattice layer at a time. By careful control of impurities, or dopants, and the thickness of the layers, electronic components can be tailor-made. Work on the manufacture and use of GaAs devices is burgeoning, especially because of the fast growth of optical laser communication technology.

•< *Gallium arsenide (GaAs) is now being manufactured as an alternative to silicon for microchips. This combination of elements is a semiconductor like silicon, but is electronically faster and therefore better for microprocessors. Silicon is a group 14 element, with four valency (bonding) electrons, whereas gallium is in group 13 with three electrons, and arsenic group 15 with five electrons.*

• T *Making microchips demands extreme control over chemical components. The layers of conducting and insulating materials that are laid down on the surface of a silicon chip may be only a few atoms thick yet must perform to the highest specifications. Great care has to be taken in their manufacture (below), and each chip is checked by test probes to ensure it performs correctly (right).*



Studying the Nucleus

The atom and the nucleus... The discovery of the nucleus...Protons and neutrons...Isotopes...Stable and unstable nuclei... The force holding the nucleus together ...Radioactivity and nuclear decay...Descriptions of the nucleus...PERSPECTiVE... The discovery of radioactivity-Rutherford and the discovery of the nucleus...Units of mass...Exotic radioactive decay

E

One of the most remarkable discoveries of the 20th century is that the apparently "solid" matter of the everyday world is actually mainly empty space. Matter consists of atoms, and each atom consists of electrons whirling round a nucleus (| page 68). The electrons endow the atoms of each element with a unique character, and determine how one atom interacts with others. Yet the electrons represent only a tiny portion - typically around 0-001 percent - of the matter within the atom. Most of the mass of an atom is concentrated in the tiny, dense nucleus; if a pea were as dense as an atomic nucleus, it would have a mass of some 10 million tonnes.

Most nuclei contain two kinds of particle - positively-charged protons and electrically neutral neutrons. (The exception here is hydrogen, the lightest element, which contains a single proton in its nucleus). The positive charge on a proton precisely balances the negative charge on an electron, and the number of protons in an atomic nucleus equals the number electrons in the atom, making atoms electrically neutral overall.

It might be thought that the electrical forces between the densely packed protons would cause them to repel one another and blow the nucleus apart. That this does not occur is due to the strong nuclear force O page 106). This operates only within the confines of the nucleus. The strong force is some 100 times as powerful as the electrical force that would otherwise disrupt the nucleus. The strong force acts equally on protons and neutrons - it does not recognize their differing electric charges. The neutrons therefore assist in keeping the protons bound together within the nucleus.

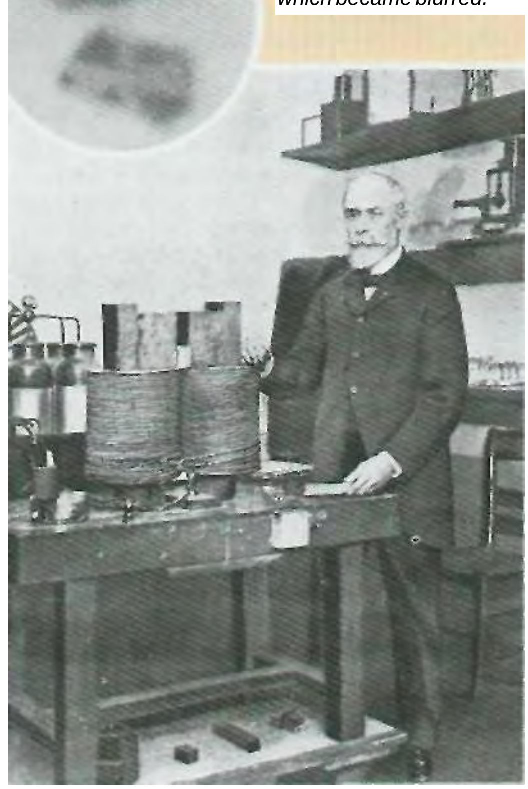
•< Atoms are typically tenths of nanometers across (10~^w m), but most of their mass lies concentrated in a central nucleus which is much smaller. The nucleus of a heavyweight atom, such as lead, is about 10,000 times smaller than the typical atomic dimension — that is, 10' "m across. To imagine the relative sizes, consider the Earth, which is about 12,750km in diameter. If an atom of lead were as big as the Earth, then its nucleus would be only about 1-3km across.

The discovery of radioactivity

In 1896, the French physicist Henri Becquerel (1852-1908) was investigating the link between fluorescence and the phenomenon of X-rays. He packed together some fluorescent uranium salt, a mask to allow through a specific pattern of radiation and a photographic plate, all of which he intended exposing to sunlight. But for several days the sun did not shine, and eventually, when he removed the package from the drawer, he discovered that an image had formed anyway. This could not be due to fluorescence but had to be due to some new form of radiation emitted by the salt.

Becquerel had discovered radioactivity, the spontaneous transmutation of an atomic nucleus from one state to another, which yields various radiations. At the time, the notion of atoms was far from being fully accepted, and it was to be many years before physicists finally came to a complete understanding of radioactivity. However, Becquerel's discovery opened the door that was to lead to the modern picture of the atom. Here, as other scientists were soon to show, was the first evidence that atoms are not the immutable objects the ancient Greeks had imagined (4 page 5). Moreover, the radiations produced by uranium and related elements were also to prove powerful tools for probing the atom, and led to the discovery of the nucleus and its contents by the British physicist Ernest Rutherford (1871-1937), and his colleagues (} page 80).

•4 f The first evidence that atoms are not immutable came in 1896 when Henri Becquerel (below) left some uranium salts next to a photographic plate, which became blurred.



Pioneering studies of the nucleus

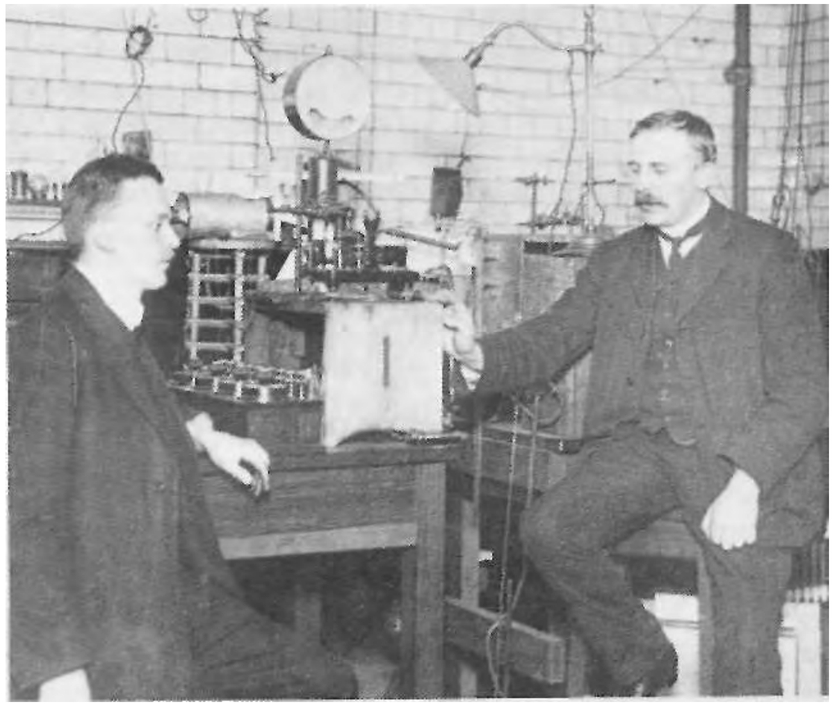
The founding father of nuclear physics was Ernest Rutherford (1871-1937). He not only discovered the existence of the atomic nucleus and the protons it contains, but also did much to elucidate the nature of radioactivity. Moreover, it was in Rutherford's laboratory at Cambridge University that the neutron was discovered.

Rutherford was born in New Zealand and studied at Canterbury College, Christchurch and Cambridge University. There he worked with the British physicist John Joseph U.J.) Thomson (1856-1940), who discovered the electron in 1897 (fpage 68). Rutherford studied radioactivity soon after its discovery in 1896, and he found two types of radiation, which he called alpha and beta.

In 1898, Rutherford moved to Montreal, where he continued his studies of radioactivity, ably assisted by a young British chemist, Frederick Soddy (1877-1956). Together they discovered how radioactivity transmutes one element to another-a natural alchemy. Rutherford later moved back to Britain in 1907, this time to Manchester University. There he worked with the young German physicist, Hans Geiger (1882-1945), and showed in 1908 that alpha particles have the same charge and mass as helium atoms. Soon after this, Rutherford encouraged Geiger and Ernest Marsden (1889-1970), a student from Blackburn, England, to study the scattering of alpha particles from gold foil. Marsden made the remarkable discovery that the alpha particles can be turned back in their tracks by the foil, and this led Rutherford to conclude in 1911 that atoms must contain a dense central nucleus.

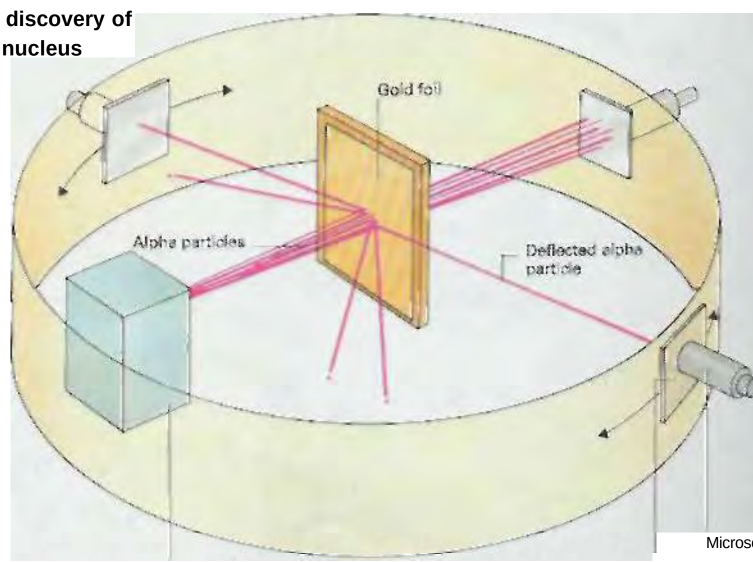
In further experiments, initiated by Marsden, Rutherford discovered that alpha particles could knock out hydrogen nuclei from a variety of other nuclei. This led him to conclude that hydrogen nuclei are basic components of all atomic nuclei, and he gave them the name of "protons ", after the Greek for "first". In 1919, around the time he published his work on protons, Rutherford succeeded Thomson as Cavendish professor at Cambridge. Thereafter he guided the efforts of a remarkable team that grew up around him, making Cambridge the world's most important center for research in experimental nuclear physics in the 1920s and 30s. One of many results that came out of the Cavendish Laboratory in a single year, 1932, was the discovery of the neutron by the British physicist James Chadwick (1891-1974).

Rutherford was a remarkable personality, renowned for his booming voice and distaste for too much theorizing. He received the Nobel Prize for chemistry for his work on radioactivity in 1908, and was made Baron Rutherford of Nelson in 1931. His ashes were placed in Westminster Abbey near the tomb of Isaac Newton (1642-1727).



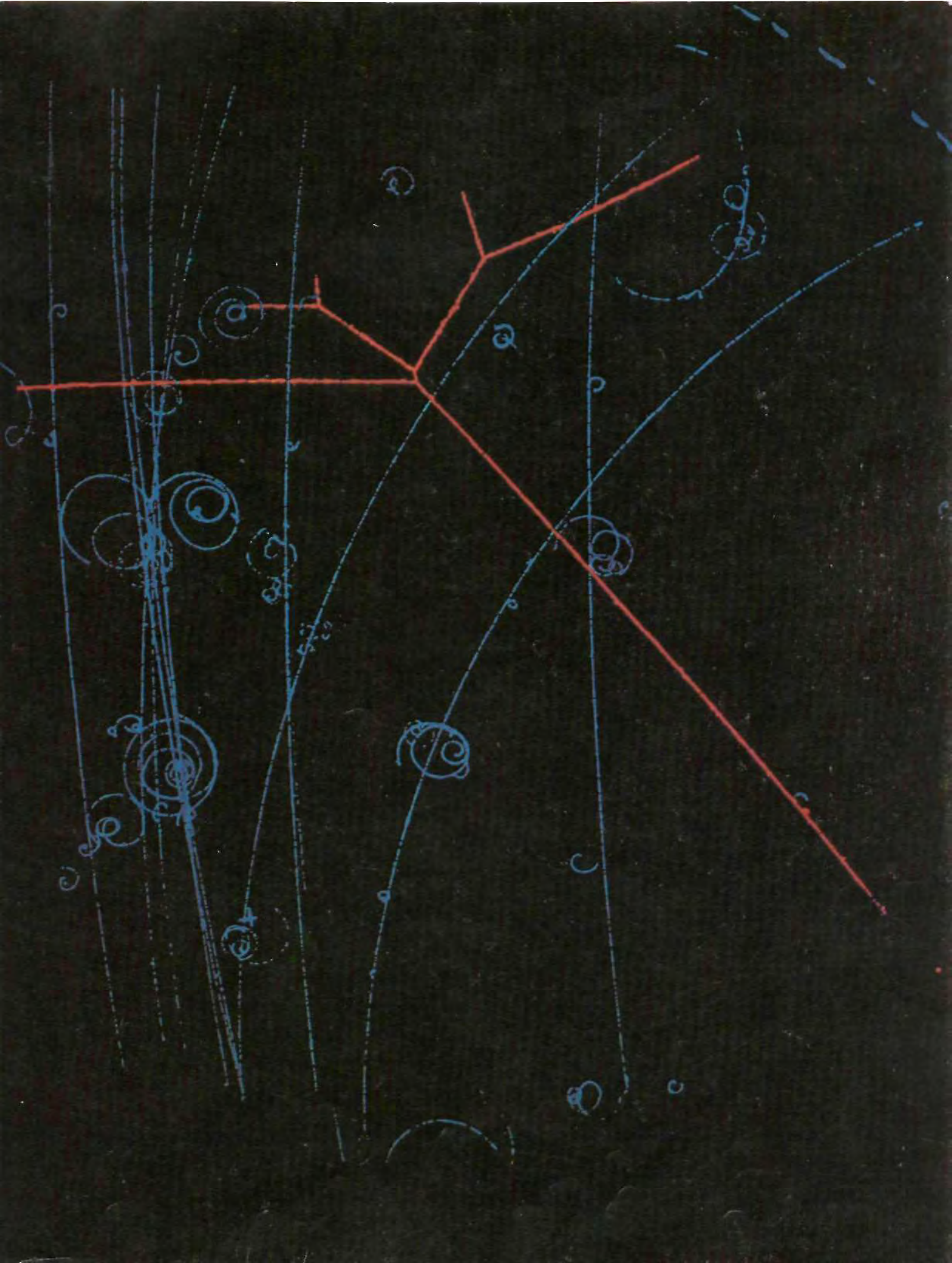
•4 A • Ernest Rutherford (above right with Geiger) masterminded both the discovery of the atomic nucleus and, within a decade, the protons that the nucleus contains. The apparatus (leftI was used in his work. Rutherford proposed that the hydrogen nucleus - the proton - must be a basic constituent in all nuclei. That protons and hydrogen nuclei are identical is shown opposite, where a proton (red) knocks many hydrogen nuclei into motion. The 90° angles between the tracks shows the equivalence of the particles.

The discovery of the nucleus



Microscope
Zinc sulfide screen
Radium source in lead box

**• In 1909, Rutherford suggested that Ernest Marsden should investigate whether alpha particles can be deflected through large angles in scattering from a thin foil. Marsden's apparatus consisted of a radioactive source emitting alpha particles, a sheet of gold foil, and a scintillating screen and microscope to view the flashes of light each time an alpha particle struck the scintillator. The screen and microscope could move through 360°. To everyone's amazement, alphas were deflected through all angles, 1 in 200,000 right back to strike the screen when positioned next to the source.*



Exotic radioactivity

Naturally occurring radioactive isotopes decay through either beta emission or alpha emission. The same is true of the many radioactive isotopes that have been made artificially, for example by firing energetic neutrons at nuclei in a target material to form new isotopes. Artificial radioactivity was discovered in 1933 by the French physicists Frederic Joliot (1900-1958) and his French wife Irene Curie (1897-1956).

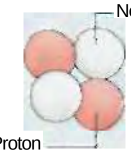
During the 1980s, however, physicists began to create exotic isotopes with either far too many neutrons or far too few. Isotopes such as the neutron-deficient lutetium-151 (71 protons and only 80 neutrons) are made in the collisions of naturally-occurring heavy nuclei, while the neutron-rich lithium-11 (three protons and eight neutrons) is one of the exotic fragments produced when uranium is bombarded with energetic protons. Such isotopes have revealed new forms of radioactivity which are prohibited to the more usual isotopes on energy grounds. Lithium-11, indeed, exhibits at least six different decay modes. It can undergo a beta decay followed rapidly by the emission of one, two, or three neutrons; or an alpha particle and a neutron; or even a triton (the nucleus ofhydrogen-3 which contains one proton and two neutrons). In a similar way, lutetium-151 can emit protons. One of the most exotic forms of radioactivity was found in 1984 in radium, one of the first radioactive elements to be discovered, nearly 100 years previously. Hadium-223, which occurs naturally in the decay chain of uranium-235, can emit relatively large clusters of protons and neutrons in the form ofcarbon-14 nuclei; so too can radium-222 and radium-224. The carbon-14 emission is very rare, occurring once for every 10 billion alpha decays.

Atomic nuclei

1

A Hydrogen is the lightest of all chemical elements, and its atomic nucleus is the simplest. It consists of a single positively-charged particle -the proton.

Neutron

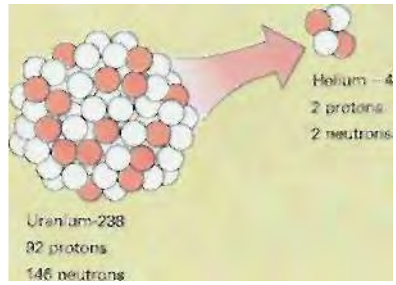


Proton

A In helium, the second lightest element, there are two protons which, being of like charge, would repel each other were it not for the two neutral neutrons.

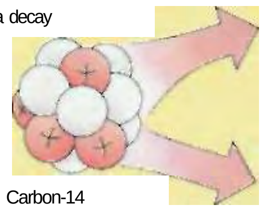
A 4s elements become heavier, both the number of protons in the nucleus increases and the number of neutrons, to overcome the electrical repulsion.

Radioactive decay



Beta decay

Electron (- charge)



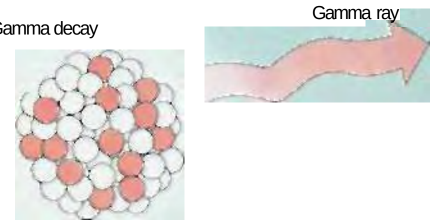
O

Neutrino (no charge)

Carbon-14
6 protons I* charge)
8 neutrons (no charge)

Gamma decay

Gamma ray



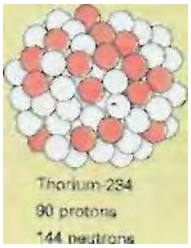
Polonium-212
"Excited" state
84 protons
128 neutrons

• Marie Curie and her daughter Irene (far right). Marie discovered the natural radioactive elements radium and polonium in the late 1890s. Nearly 40 years later, Irene and her husband Frederic Joliot discovered artificial radioactivity; they observed that aluminum brought into contact with polonium became radioactive, emitting positrons (positive electrons) continually for several minutes after the polonium was removed. Alpha particles from the polonium had interacted with nuclei in the aluminum to produce a radioactive form of phosphorus. The radioactive phosphorus emitted positrons, turning into a stable form of tin.

< Now, machines such as the Bevatron at Berkeley produce high energy nuclei to create exotic forms of radioactivity.

A

Proton O Neutron O Electron

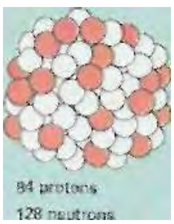


M In alpha decay, a heavy nucleus such as uranium-238 changes toward a more stable form by spitting out an alpha particle — two protons and two neutrons la helium nucleus). Because the total number of protons changes, the type of chemical element changes.

Nitrogen 14
7 protons
7 neutrons

•4 In beta decay, a nucleus with too many neutrons becomes more stable when one of the neutrons changes into a proton, emitting an electron and a neutrino. Unlike alpha decay, the overall number of neutrons and protons remains the same.

Polonium-212
"Ground" state



•4 Nuclei formed in various nuclear reactions, including alpha and beta decay, can be created with too much energy: they are "excited". Such nuclei lose this excess energy by emitting a gamma-ray - a form of high energy electromagnetic radiation.

The energy of a nucleus

Certain configurations of protons and neutrons have less energy than others, and the most stable configurations are those with least energy. This is like observing that a ball rolls down a hill to the bottom because its gravitational potential energy (| page 18) is less at the bottom of the hill. The "energy" of a nucleus is a subtle concept to grasp; it is in fact the mass of the nucleus. Albert Einstein showed in his special theory of relativity (^ page 42) that mass is a form of energy. Thus a nucleus with a minimum energy has a minimum mass.

The mass of a stable nucleus is always less than the total mass of the protons and neutrons it contains. For example, the mass of nilrogen-14, the commonest isotope of nitrogen, is less than the total mass of the seven protons and seven neutrons it contains. However, the mass of nitrogen-14 is also less than that of carbon-14 - the radioactive form of carbon used in carbon dating. Carbon-14 also contains 14 nuclcons (eight neutrons and six protons), but its mass-energy is 0.0012 percent greater than that of nitrogen-14. Carbon-14 decays to nitrogen-14, shedding its excess mass-energy as it does so.

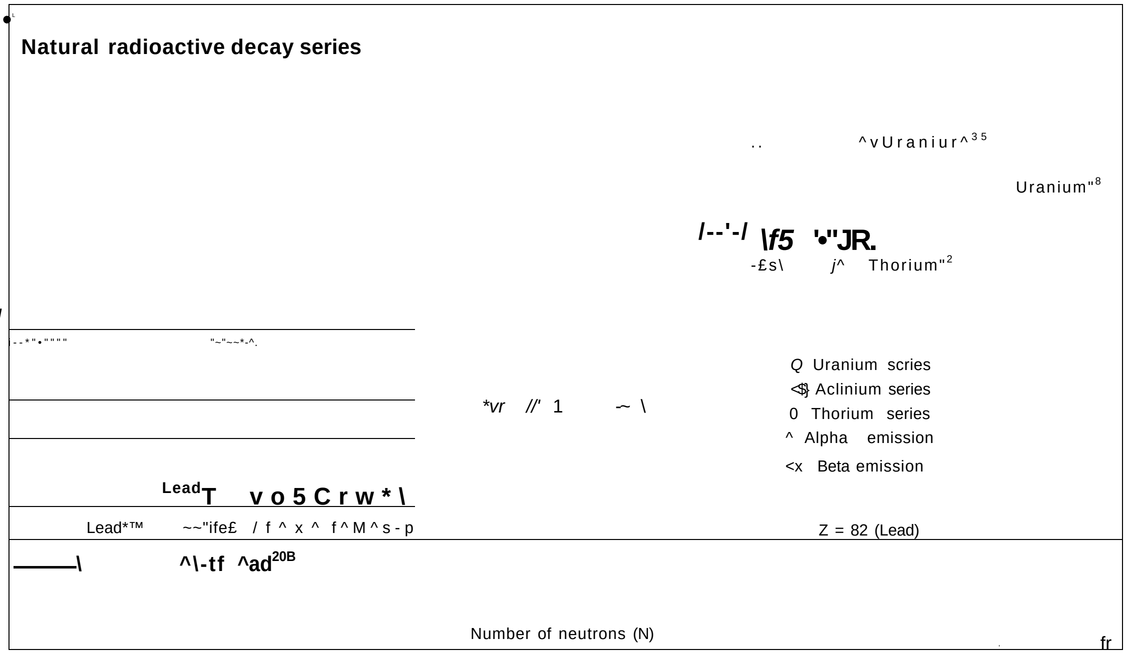
Radioactive decay

The way in which unstable nuclei rid themselves of their excess energy depends in part on how much energy they have over and above a more stable nucleus. Radioactivity can occur in a variety of ways. The three most common forms are known as alpha, beta and gamma decay. In alpha decay, a nucleus spits out an alpha particle. This is a conglomeration of two protons and two neutrons (in other words, a helium nucleus). To emit an alpha particle, a nucleus must have a larger mass-energy than the total mass of the new nucleus and the alpha particle. This is often true for nuclei heavier than lead, which can lose energy by emitting an alpha particle, but which would need additional energy to give up a single proton or neutron.

In beta decay, the nucleus ejects a different kind of particle - an eleclron. However, unlike the alpha particle whose protons and neutrons inhabited the initial nucleus, the electron did not exist inside the nucleus before emerging. It is created at the moment the nucleus decays, as a neutron transmutes into a proton. To do this, the nucleus must rid itself of some energy and increase its electric charge by one unit. The electron, with its negative charge, ensures the appropriate change of charge, and it also carries away a small fraction of the excess energy as its own mass. The balance of energy for the whole process, released as the nucleus converts from one kind to another, is, however, shared between the motion of a third product of the decay - a neutral particle called the neutrino (^ page 102).

Beta decays of this kind occur in nuclei with too many neutrons. This is how carbon-14 decays to nilrogen-14. Nuclei that are too rich in protons can decay by another form of beta radioactivity. In this case, a proton converts into a neutron by emitting a positron along with a neutrino. The positron is in fact a positively charged version of the electron - it is an antielectron (» page 101).

The third common form of radioactivity is gamma decay. The nucleus does not emit a particle but a very energetic form of electro-magnetic radiation (| page 61). Gamma radiation does not change the number of protons or neutrons in a nucleus, but it does reduce the energy of a nucleus. It occurs when a nucleus has changed through alpha or beta radioactivity to an "excited" nucleus, in which the protons and neutrons jostle each other with more energy than usual.



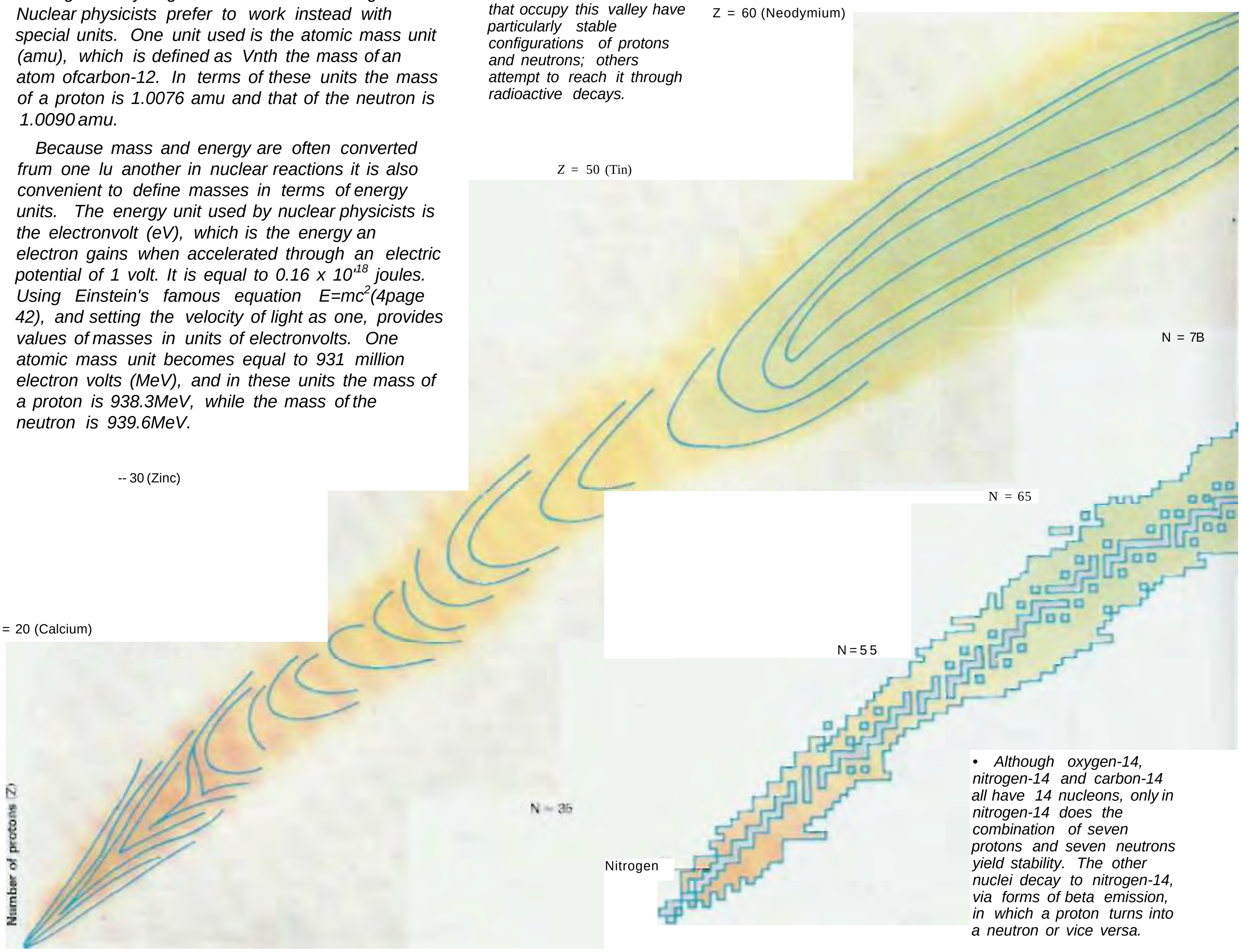
< At the top end of the valley of stability are found the heaviest nuclei, all of which are unstable. Through emitting alpha or beta radiation, these nuclei can follow zigzag paths that take them to stable configurations of the element lead. Three such pathways allow the decays of naturally-occurring radioactive nuclei.

The units of atomic mass

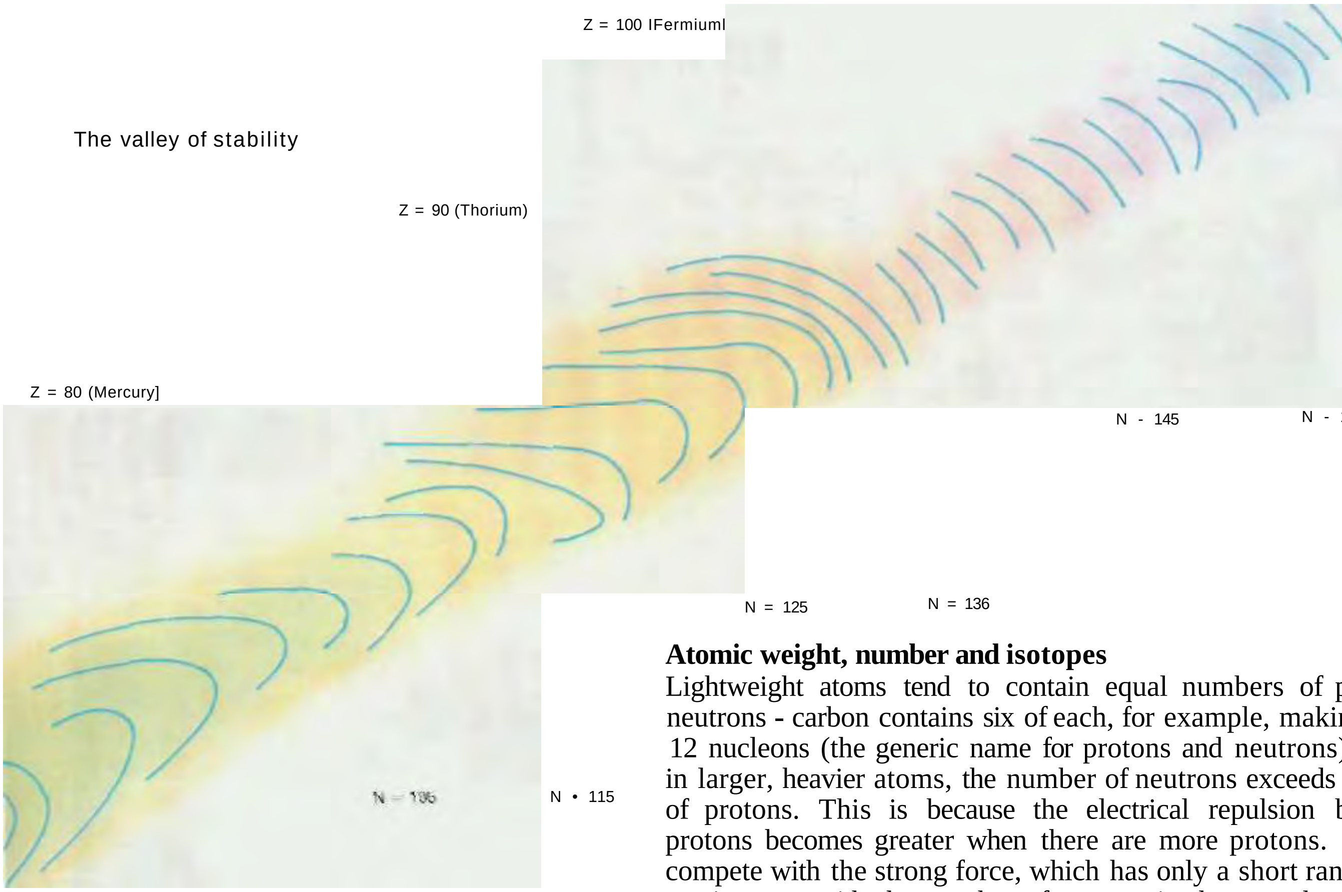
Conventional units of mass prove unwieldy when dealing with atomic nuclei. The mass of a proton is 1.673 x 10^{-27}kg, while that of a neutron is slightly -but significantly-larger at 1.675 x 10^{-27}kg. Nuclear physicists prefer to work instead with special units. One unit used is the atomic mass unit (amu), which is defined as Vnth the mass of an atom ofcarbon-12. In terms of these units the mass of a proton is 1.0076 amu and that of the neutron is 1.0090 amu.

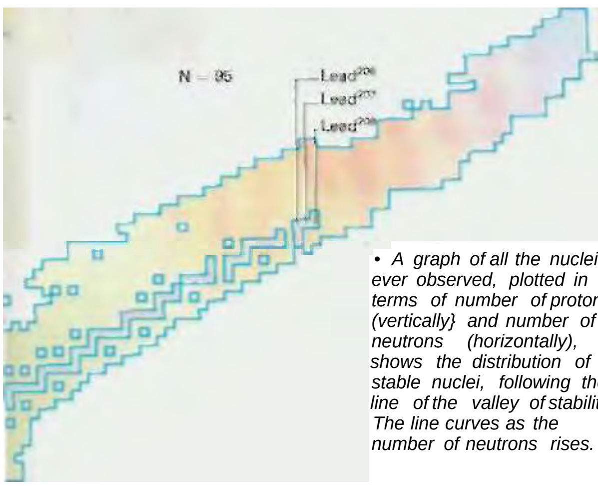
Because mass and energy are often converted frum one lu another in nuclear reactions it is also convenient to define masses in terms of energy units. The energy unit used by nuclear physicists is the electronvolt (eV), which is the energy an electron gains when accelerated through an electric potential of 1 volt. It is equal to 0.16 x 10^{-18} joules. Using Einstein's famous equation $E=mc^2$ (4page 42), and setting the velocity of light as one, provides values of masses in units of electronvolts. One atomic mass unit becomes equal to 931 million electron volts (MeV), and in these units the mass of a proton is 938.3MeV, while the mass of the neutron is 939.6MeV.

• A "contour map" of nuclear energies for a wide range of nuclei shows a "valley of stability" where the energies have their minimum values. The nuclei that occupy this valley have particularly stable configurations of protons and neutrons; others attempt to reach it through radioactive decays.

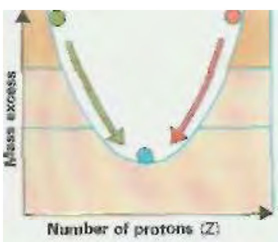


• Although oxygen-14, nitrogen-14 and carbon-14 all have 14 nucleons, only in nitrogen-14 does the combination of seven protons and seven neutrons yield stability. The other nuclei decay to nitrogen-14, via forms of beta emission, in which a proton turns into a neutron or vice versa.





• A graph of all the nuclei ever observed, plotted in terms of number of protons (vertically) and number of neutrons (horizontally), shows the distribution of stable nuclei, following the line of the valley of stability. The line curves as the number of neutrons rises.



•4 A section through the valley of stability for nuclei with the same total number of protons and neutrons shows how the nuclear energy (or mass excess) varies with the number of protons, Z. The value of Z at the valley bottom gives the preferred stability value.

Atomic weight, number and isotopes

Lightweight atoms tend to contain equal numbers of protons and neutrons - carbon contains six of each, for example, making a total of 12 nucleons (the generic name for protons and neutrons). However, in larger, heavier atoms, the number of neutrons exceeds the number of protons. This is because the electrical repulsion between the protons becomes greater when there are more protons. It begins to compete with the strong force, which has only a short range and does not increase with the number of protons in the way that the electric-force does. The effect of extra neutrons in a larger nucleus is to dilute the electrical repulsion, so that the nucleus remains bound together. Uranium, for example, contains 92 protons and 246 neutrons. In nuclei not much larger than this even the addition of extra neutrons can no longer prevent the nucleus from blowing itself apart. Although some nuclei with more than 100 protons can be made, they arc very unstable and exist at most for only a few hours.

The number of protons in a nucleus is known as the atomic number, Z, and it is unique for each element; thus carbon has Z=6, oxygen Z=8 and so on. However, the number of neutrons can vary without influencing the chemical properties of an atom, which depend only on the number of electrons and hence on the atomic number. Zinc, for example, occurs naturally in a form containing 30 protons and 34 neutrons. But only just under 50 percent of zinc is in this form. About 28 percent of zinc nuclei contain a total of 36 neutrons, while nearly 19 percent contain 38 neutrons. The remaining 3 percent contain still different numbers of neutrons. These different nuclei are called "isotopes"; this word derives from the Greek for "the same place", because they have the same value of Z and therefore occur in the same place in the table of elements. To differentiate isotopes, they are referred to by the total number of neutrons and protons they contain. Thus natural zinc contains a combination of the isotopes zinc-64, zinc-66 and zinc-68.

All elements have a range of possible isotopes, although often only one isotope dominates, forming 90 percent or more of the substance. This is because all isotopes other than the most common ones are intrinsically unstable, and transform into a different kind of nucleus after a period of time that can vary from fractions of a second to millions of years. These are called radioactive isotopes because they change through one of the several processes that are together known as radioactivity.

See also
Atoms and Elements 65-72
The Quantum World87-96
Elementary Particles 97-104
Radiation and Radioactivity 111-18
Nuclear Fission and Fusion 119-24

Theories of the nucleus

Developing a theoretical model that explains all the observed properties of stable, unstable and excited nuclei has proved a difficult challenge. One of the first models to achieve any degree of success was the "liquid drop model", first developed by two theorists, the Dane Niels Bohr (1885-1962) and the American John Wheeler (b.1911) in 1939. This draws analogies between certain properties of the nucleus and those of a drop of liquid, and it proves particularly useful in discussing the "fission" of heavy nuclei into two fragments (•page 121).

However, the liquid drop model cannot deal with one important property of nuclei, namely "spin". Quantum theory shows that protons and neutrons behave as if they possess an intrinsic angular momentum, rather like subatomic spinning tops (> page 93). A full theory of the nucleus must take into account the spins of the constituent nuclcons, and predict how they add together to give an overall spin angular momentum for the nucleus as a whole. One model based on spin is the "shell model", developed independently in 1948 by physicists Maria Goeppert-Mayer (1906-1972), in the United States, and Hans Jensen (1907-1973), in Germany, both of whom received the Nobel Prize for physics in 1963.

In this model protons and neutrons occupy "shells" analogous to the shells of atomic theory (| page 69). One of the driving forces behind such a model is the observation of the so-called magic numbers. Experiments indicate that certain numbers of protons (Z) and neutrons (N) give nuclei that are exceptionally tightly bound. In particular the values 2, 8, 20, 28, 50, 82, 126 are associated with this stability. For example, oxygen-16 (Z=8, N=8), calcium-40 (Z=20, N=20), and lead-208 (Z=82, N=126) are very stable; and tin, with Z=50, has as many as 10 stable isotopes. According to the shell model the magic numbers are associated with closed shells of nucleons.

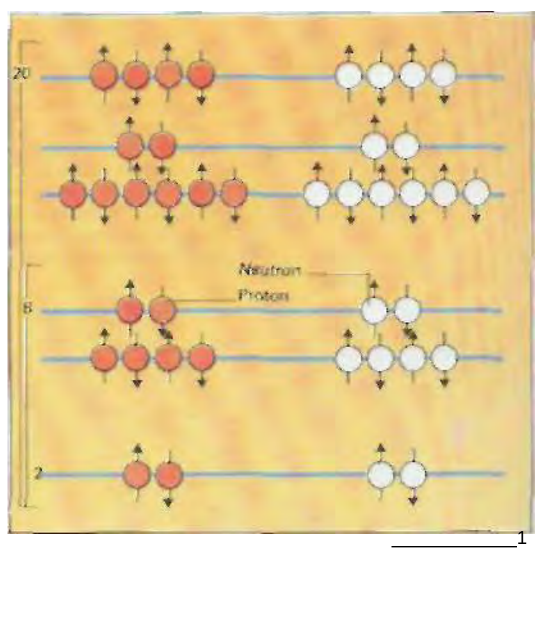
The collective model of the nucleus

The shell model proves successful in accounting for the observed spins of nuclei, but it runs into difficulty with other nuclear properties. It fails particularly in accounting for non-spherical distributions of electric charge that occur in many nuclei. To deal with problems of this kind, Aage Bohr (b.1922), a Danish theorist and son of Niels Bohr, and Ben Mottleson (b.1926), a theorist in the United States, developed the "collective model" of the nucleus. This had been first proposed in 1950 by another American physicist, James Rainwater (b.1917). The collective model attempts to unite the best features of the shell model and the liquid drop model. It does so by regarding a large nucleus as a closed-shell core surrounded by additional "valence" nucleons. This core can be deformed from its spherical shape by the interactions of the outer nucleons. This model has particular success in describing the vibrations and rotations of atomic nuclei which can be studied experimentally.

A general theoretical description that can be applied to all nuclei remains the goal of nuclear physicists. Data for these studies comes from experiments at machines that can accelerate ions of heavy elements - up to uranium in some cases - and collide them with targets to create "exotic" isotopes with unusual numbers of protons and neutrons, which are sometimes also spinning very rapidly. These isotopes can be greatly deformed, and they provide stringent tests of the theoretical models.

A *Niels Bohr (right) and his son Aage both made great contributions to the theory of nuclear structure. Niels, with John Wheeler, put forward the liquid drop model, while Aage has attempted to unify this with the shell model.*

A T *Maria Goeppert-Mayer (left) with Hans Jensen. They independently developed a "shell" model for nuclei in the late 1940s, in which protons and neutrons fit into "shells" within a nucleus. The simplest shells can contain two protons and two neutrons — provided the members of each pair spin in opposite directions. In other shells, the nucleons have additional angular momentum and more particles are allowed. Full shells, corresponding to a total of two, eight, or twenty neutrons or protons, for example, yield very stable nuclei. The diagram illustrates the shell allocations for calcium-40, which has closed shells both of protons and neutrons, with 20 particles in each.*



The Quantum World

12

The concept of quanta... The particle theory of light... Planck's constant...Bohr's image of the atom...Particles as waves...Quantum mechanics...PERSPEcnvE... Theories of light emission...Niels Bohr... The photoelectric effect...Lasers...Electron microscopy... Spin and NMR...Philosophical implications of the quantum theory

Many natural physical phenomena are distinguished by their continuity. The Earth smoothly orbits the Sun; the ground slowly warms on a sunny day, only to release its heat in an equally gentle fashion during the night. But in the world within the atom, the story is very different. There, subatomic systems can exist only in certain fixed states. In other words, many subatomic phenomena are "quantized". It is like comparing stairs with a smooth slope - whereas a person can stand at any height on a slope, only fixed "discrete" heights are possible on the staircase.

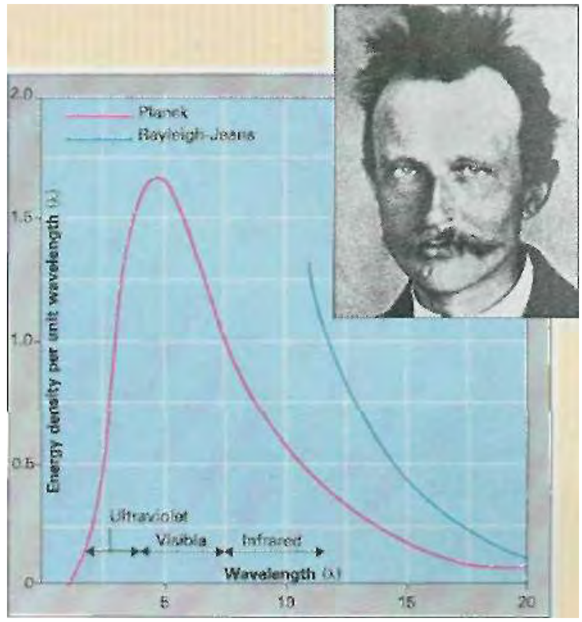
Quantization is one important feature of subatomic physics. Another is its probabilistic nature. An astronomer can say with certainty where the Earth, Moon and the other planets will be in relation to the Sun on a given date, but an atomic physicist can state only the *probability* that an electron will be in a given state at a given time. In addition, there is a fundamental limit to the precision of statements about the subatomic world. The more precisely the value of one property is known, the less precisely the value of a second related property can be known (| page 96). Such features make the subatomic world very unfamiliar from an "everyday" viewpoint.

The particle theory of light

One prime example of quantum physics at work is in the behavior of light. Physicists in the 19th century found good evidence that light is a wave motion (| page 44), a view reinforced in the electromagnetic theory of the British physicist James Clerk Maxwell (1831-1879) in 1865. Maxwell's theory predicted the existence of electromagnetic waves traveling through free space at a specific velocity, close to the measured velocity of light. This showed that light is just one variety of a whole range of electromagnetic radiation with varying wavelengths, all of which travel through free space at the same velocity.

The wave theory works well in describing the movement of light through different materials. It explains the phenomena of reflection, refraction, and the diffraction and interference of light (| pages 38-39). But the theory breaks down when it comes to explaining the absorption and emission of light.

The first person to come to terms with these difficulties was the German physicist Max Planck (1858-1947). Physicists in the late 1890s were faced with a problem in that they could not explain the way that the intensity of radiation emanating from a "perfect emitter" varies with wavelength. There is a peak in intensity at a wavelength associated with the temperature of the emitting body: the higher the temperature, the shorter is this maximum wavelength. But in the late 19th century, theoretical understanding implied that the intensity should go on rising at ever shorter wavelengths, rather than falling down from a peak. There was clearly something wrong with the theory.



A Max Planck developed his quantum theory of radiation in an attempt to resolve a difficulty with the existing theory for the emission of radiation. In the theory of Rayleigh and Jeans, the amount of energy emitted increased limitlessly at shorter wavelengths. This was an "ultraviolet catastrophe" that was clearly at odds with experimental results.

Understanding the emission of light

In the late 19th century Heinrich Hertz demonstrated that light is an electromagnetic wave It page 60). At the same time, observations were made that were difficult to reconcile with Maxwell's theory, such as the emission of light at only a few wavelengths from certain sources. Such sources yield a spectrum of discrete lines when their light is passed through a prism. Even continuous spectra provided a problem, when physicists tried to explain how intensity varies with wavelength.

The British physicists Lord Rayleigh (born John Strutt, 1842-1919) and James Jeans tried to explain the spectrum in the following way. They considered a perfect emitter (or "black body") as a box which can store electromagnetic energy as "standing " waves, like the sound waves in an organ pipe. The spectrum of waves emerging from a hole in the box should, they argued, correspond to the observed emission spectrum from a source like the Sun. They could calculate the number of standing waves within a particular range of wavelengths. And they assumed that the same amount of energy was associated with each wave.

The Rayleigh-Jeans law agreed with observation at high wavelengths, but disagreed at lower wavelengths—indeed, it implied that an infinite amount of energy could be radiated at low wavelengths. Planck discovered the flaw in the theory. The concept of standing waves in a box and the number of waves were not at fault; it was the assumption that the energy is distributed equally between different wavelengths. Planck associated different energies with different wavelengths. He thereby resolved the problem, and gave birth to quantum theory which was eventually to explain the emission lines of other types of source.

Einstein received the Nobel Prize for his work on the photoelectric effect, rather than his work on special relativity

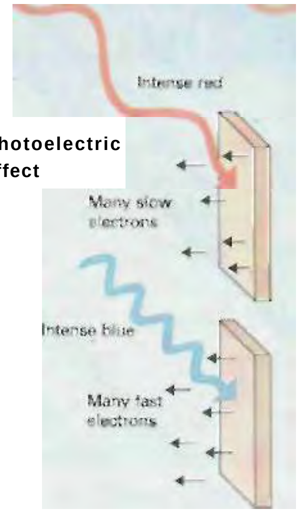
Planck was able to solve the problem of light emission by postulating that the energy in radiation is emitted in fixed amounts, and that the size of each "quantum" of energy depends on the frequency at which the radiation is oscillating. The energy is given by the frequency of the radiation multiplied by a constant: energy equals frequency times a constant. Known as Planck's constant, it is usually symbolized by the letter *h*. It has a value of 6·6 x 10^{-34} joule seconds. Only in systems where the relevant energies are very small, as in individual atoms, does the quantum concept become important. The equation also reveals how the energy associated with a quantum of radiation varies across the electromagnetic spectrum. Radiation at high frequencies (short wavelengths) consists of high-energy quanta; at low frequencies (long wavelengths) the quanta have low energies.

The German physicist Albert Einstein (1879-1955) used the same idea of quantized radiation to explain the photoelectric effect, in which ultraviolet light causes atoms in the surface of some metals to emit electrons. Energy of the emitted electrons depends not on the *intensity* of the ultraviolet radiation, as mighl be expected if the radiation is a wave, but on the *frequency*. Also, there is no emission below a certain frequency, characteristic of each metal. Einstein argued that the radiation must arrive at the surface in quantized "wave packets", each with an energy related to the frequency through Planck's equation. Thus, radiation at too low a frequency does not have enough energy to knock out even one electron, but at higher frequencies it knocks out electrons with increasing energies.

Light thus exhibits a dual behavior. In certain circumstances it proves convenient to regard it as a wave, in others as a stream of quanta or "particles" of light. If an electron in an atom absorbs a photon, its energy increases by the amount of energy the photon carried; if the electron emits a photon, its energy decreases. However, the energies of the electron are also quantized. Like a person standing on a staircase, it can absorb or give up energy only in fixed amounts as it moves from one energy level (stair level) to another. Thus an atomic electron cannot emit a photon of any energy.

There is a precise relationship between the situation of an electron within an atom and the energy of the photons it can absorb or emit. The allowed energy levels depend critically upon the positive charge of the nucleus about which the electrons circulate. In other words, each chemical element has its own individual "staircase" of energy levels *(i* page 68). This in turn implies that the photons emitted (or absorbed) by atoms of different elements have different energies (and wavelengths), which correspond to the different gaps in each allowed series of energy levels. Each atom thus has its own unique emission and absorption spectra, which are as good an indicator of an atom's identity as fingerprints are of a human's.

Photoelectric effect



Dim red

15k

Few slow efeclfons

<**%

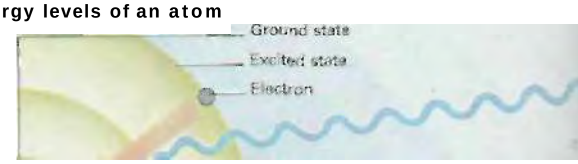
Dim blue

r^

Few fast electrons

A In the photoelectric effect, light knocks electrons out of a material. It can be explained only if light consists of photons, localized wave "packets", as opposed to the continuous waves of classical electromagnetic theory. Changing the intensity of the light (top) changes the number of photons: this alters the number of electrons emitted, but has little effect on their energy. Changing the frequency of the light, however, changes the energy of the photons and therefore the energy of the electrons emitted. In practice, high frequency, ultraviolet light is often necessary.

Energy levels of an atom



Nucleus

O

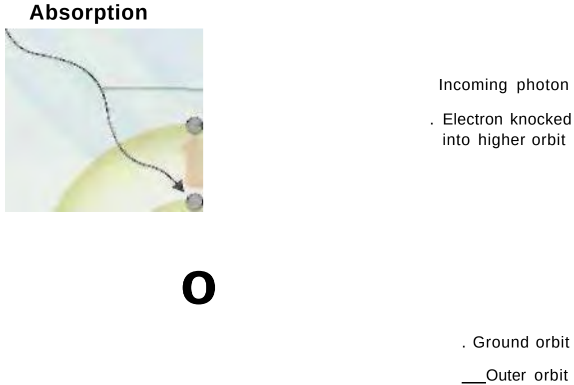
A /Electrons in atoms emit light when they change energy, often after being excited to higher energy levels than usual, perhaps by heating. With each jump from a higher energy level to a lower one, an electron emits a single photon. The energy (frequency) of the photon depends on the size of the jump: thebiggerthe jump, the greater the energy of the emitted photon, and the higher its frequency.

f The frequencies of light emitted or absorbed by atoms provide their "fingerprints". The emission spectrum of helium (left) shows the light produced as excited electrons jump down from higher energy levels. A spectrum of the light from the Sun (right) shows dark lines where photons of particular frequencies have been absorbed as electrons in atoms jump to higher energy levels.

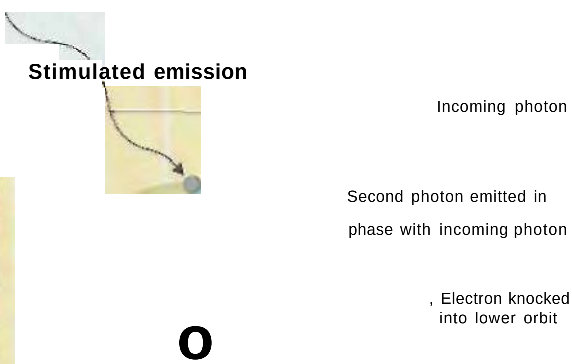
A *Lasers are key research tools. They provide a source of light that is not only monochromatic (single wavelength) but also coherent- the emitted photons are in perfect step with each other. This greatly enhances the intensity of light.*

• *When an atom absorbs light, electrons are raised to higher energy levels by photons of the appropriate energy (topi, they can then emit photons and return to their usual levels spontaneously. However, they can also be stimulated to fall back by a photon of frequency corresponding to the change in energy levels. This is stimulated emission - the emitted photon and the original incident photon move away in step.*

Absorption



Stimulated emission



Lasers

The acronym laser stands for "light amplification by stimulated emission of radiation". The crucial words here are "stimulated emission ". In ***an*** *emission process, a photon is produced when an electron falls from one atomic energy level to another, lower level. In stimulated emission, this process is encouraged by radiation of the same energy (that is, the same frequency and wavelength) as the emitted radiation. Thus one photon stimulates the emission of a second photon of exactly the same wavelength. Moreover, the two photons are precisely in step such that the peaks and troughs of the resulting radiation match exactly. The radiation produced in this way is said to be "coherent", and it is valuable for carrying information, and for the extremely high energy-density that a beam of such radiation can possess. This is the value of a laser beam.*

For stimulated emission to be useful, an atomic system needs more atoms with electrons in the relevant upper energy level than is normally ***the*** *case. In a laser, this is achieved by "optical pumping". A flashlight, for example, can "excite" many electrons into the upper level. These electrons then begin spontaneously to fall back down to the lower energy level, emitting radiation as they do so.*

The principle of a laser is to trap some of this radiation so that it stimulates further emission, and the device emits a coherent beam. This is done by mirrors to ensure that the emission occurs within a "resonant cavity", in which the radiation bounces back and forth, stimulating more emission in the desired way, before it escapes.

The wave nature of the electron is seen in diffraction patterns similar to those of light

Applying the quantum theory to the atom

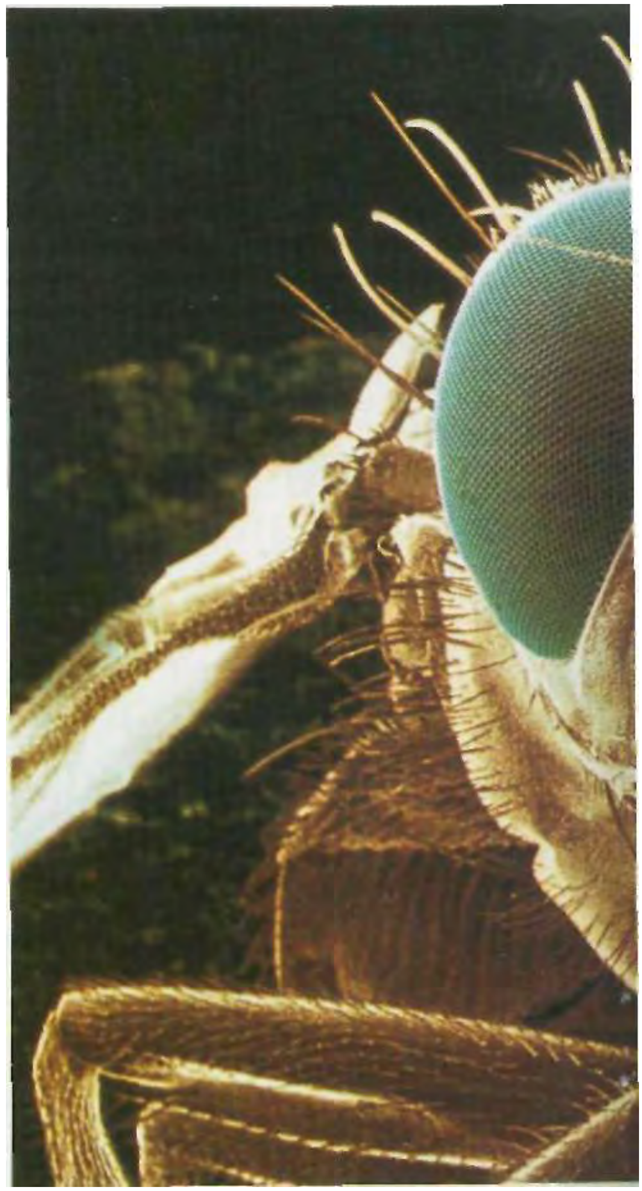
The Danish physicist Niels Bohr (1885-1962), working from Rutherford's concept of the atom as a solar system in which electrons orbit at random around the tiny nucleus, first put forward the idea of quantized atomic structure in 1913 (| page 65). In so doing he took Planck's concept a stage further by quantizing not the energy but the angular momentum (^ page 12) of the electrons in an atom. Bohr postulated that the angular momentum can take only those values equal to Planck's constant multiplied by an integer (a whole number) and divided by another constant (2 times w, the familiar geometrical constant). The angular momentum of an electron orbiting a central nucleus depends on the radius of the electron's orbit, so this quantization condition implies that only certain orbits are allowed. Bohr used it to calculate the energies of the electrons in these allowed orbits, and so worked out the staircase of energy levels in an atom.

The theory works well with the hydrogen atom, which has only one electron, but it was too simplistic for more complex atoms (| page 68). However, it could account for the wavelengths of the observed spectral fingerprints of hydrogen. Moreover, the idea of atomic energy levels allows a relatively simple understanding of physical phenomena that depend on behavior at the subatomic level - for example, the operation of lasers. And it set others on the road to developing mathematical tools for dealing with the quantum world.

Particles as waves

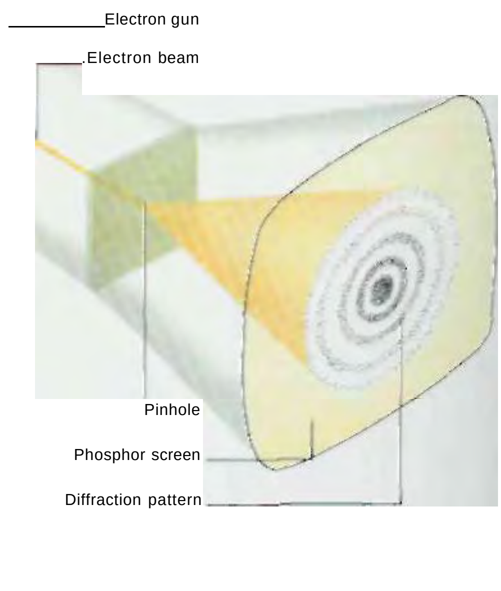
Modern understanding of the quantum nature of the atom includes yet another unusual feature of matter at the subatomic level - the wave nature of particles. In many respects the photon is like a particle of light. It has no mass, but it does have energy, and as the energy of an object is related to its momentum, it is possible to calculate the momentum of a photon. This momentum is equal to Planck's constant divided by the wavelength. In 1924, the French aristocrai and physicist, Louis de Broglie (1892-1987) suggested in his doctoral thesis that this relation could be turned around to define a wavelength for a particle such as an electron. Thus, the wavelength of the particle is equal to Planck's constant divided by its momentum. This proposal revealed for the first time the subtlety of nature that not only makes radiation behave as particles in the subatomic world, but which also makes particles behave as waves. Thus, although the energy levels of the electrons calculated by Bohr are usually depicted as orbits of increasing distance from the nucleus, they can also be seen as waves orbiting the nucleus, their frequency rising with increasing energy.

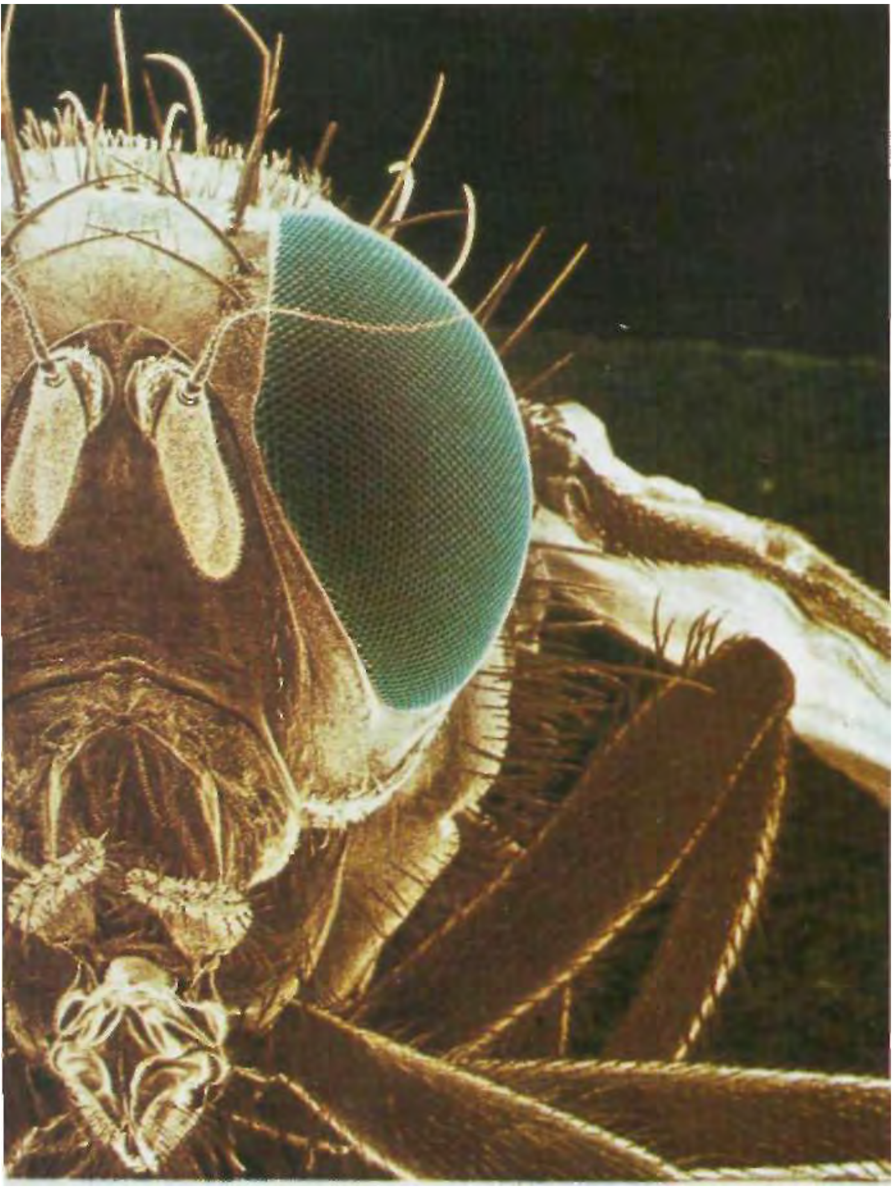
The wave nature of particles was first shown in 1927 in an experiment by two American physicists Clinton Davisson (1881-1958) and Lester Germer (1896-1971). They directed a beam of electrons at the surface of a crystal of nickel, and detected those which were scattered back from the surface. They found that the pattern of scattered electrons showed a peak at an angle of 50°, and this could be explained only if the regular array of atoms in the surface of the crystal was scattering electron *waves*. Not long afterwards, the British physicist George (G.P.) Thomson (1892-1975) observed a pattern of rings in electrons directed through a thin metal foil: such a pattern is characteristic of diffraction effects, which are well understood in terms of wave motions (• page 22). Today the diffraction of neutron beams is also used routinely for the purposes of investigating the structure and properties of materials.



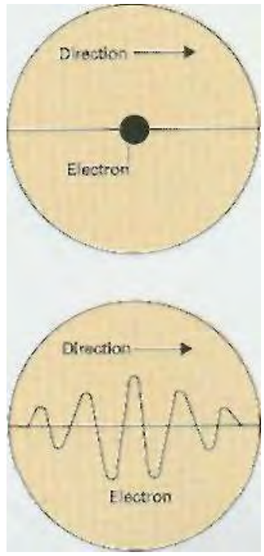
T *The wave nature of the electron can be shown by the creation of diffraction patterns similar to those of light It page 42). If a stream of electrons passes through a pinhole, the result is a spray of electrons, rather than a single point hitting the phosphor screen.*

A *A scanning electron microscope reveals detail in an image according to the number of electrons that scatter as a fine electron beam scans the specimen. The detail is possible because the wavelength of electrons is much less than that of visible light.*





• The particle description of the electron regards it as an object at a point in space and time with a definite velocity. But the concept of an electron as a wave is also valid. In this case the electron is described by a "wave function" which covers an extended region of space. The square of the wave function gives the probability of finding the electron at a given point in space and time. The velocity of this wave "packet" is the same as the velocity in the particle description. But in the wave description the precise position and momentum of the electron cannot be simultaneously determined - this is the uncertainty principle (t page 96).



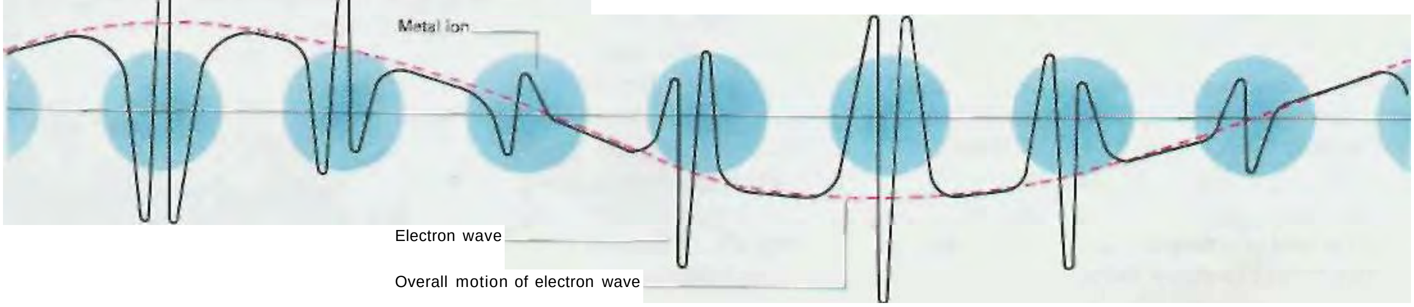
Bohr and the atom

Niels Bohr had only recently received his doctorate from the University of Copenhagen when he went to work at Cambridge University in 1911. There he encountered the New Zealand physicist Ernest Rutherford 11871-1937) while on a visit to Cambridge, and Bohr was so impressed that he soon moved to Manchester University to work with Rutherford. Rutherford had recently discovered that most of the mass of an atom is concentrated in a central dense nucleus I * page 80) and he advanced the concept of the atom as a miniature "solar system" with electrons orbiting the nucleus. One problem with this picture, however, was that according to electromagnetic theory (4 page 60) the electrons should radiate energy as they whirl round the nucleus, and therefore should spiral inwards towards the center of the atom as they lose energy.

Bohr set about resolving the problem with the aid of the new ideas of quantization. He postulated that an electron does not radiate continuously as "classical" theory demands, but that it emits energy only when it moves from one fixed orbit to another. However, Bohr did not quantize the energy of the electrons in fundamental units. Instead he • quantized their angular momentum. In so doing he broadened Planck's concept, which had applied to the energy of radiation, and showed that in the subatomic world other basic quantities can be quantized. He was awarded the Nobel Prize for physics in 1922 for his work on quantum theory, and became one of the most influential physicists of the century, developing the philosophical understanding of quantum theory.

•* Davisson and Germer were the first to show the diffraction of electrons and hence their wavelike nature. Davisson is holding the electron tube they used.

T Classical physics found it hard to explain how electrons flow easily along a metal wire when conducting electricity If page 52). The waveform of an electron moving through the lattice of ions in a metal crystal is modified close to the ions. The net effect is to allow the electron to move easily, without colliding many times.

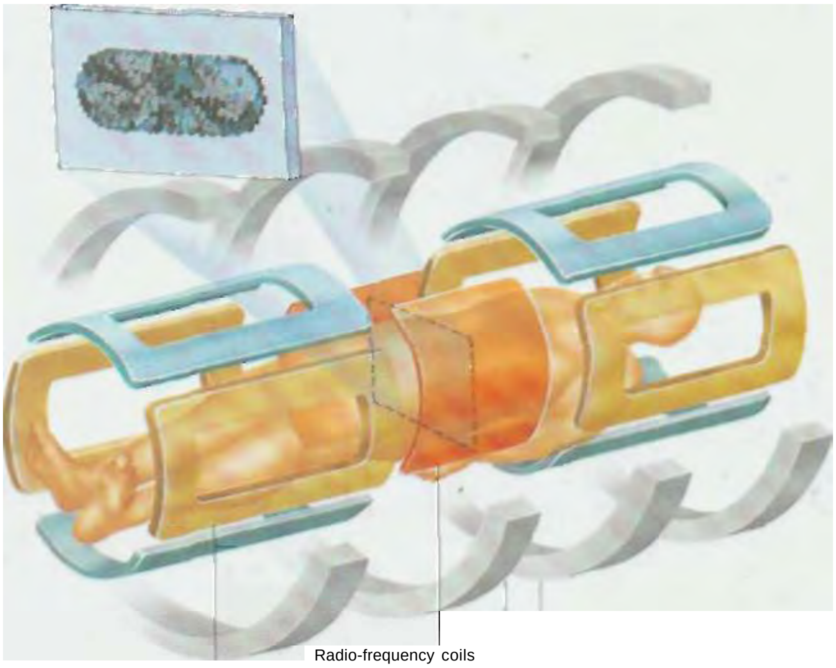


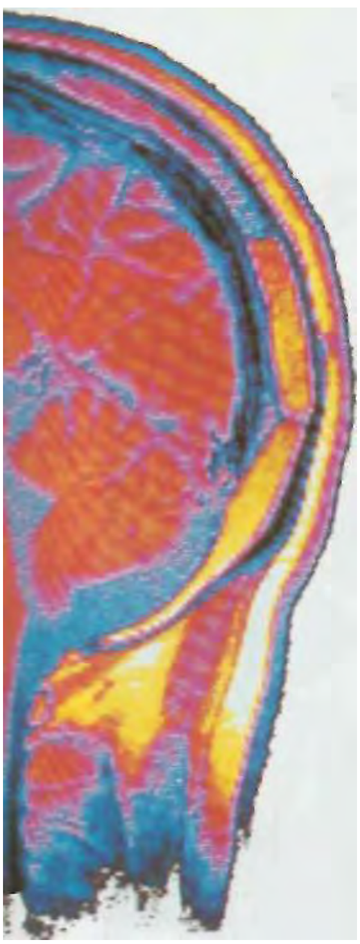
Wave mechanics

The discovery of the wave nature of particles provided a deeper understanding of the meaning of Bohr's momentum quantization. The quantization condition allows only those orbits into which a whole number of electron wavelengths fit. With this realization, and the representation of electrons as wave "packets", it is possible to develop a theory of "wave mechanics" that describes not only the structure of atoms, but many aspects of subatomic behavior.

The key feature of wave mechanics is that a particle like an electron is represented mathematically by a "wave function", which is a measure of the probability that the particle is in a given state. In the atom, Bohr's fixed orbits are replaced by "orbitals" - regions of space in which the probability wave of an electron can fit (4 page 69). Thus the orbital does not give a precise location for an electron; rather it gives the probability that an electron is more likely to be remote from the atomic nucleus, for example, than close to it. This modern quantum theory of the atom may sound very "woolly", yet it provides a surprisingly cast-iron framework for much of what scientists today know about physics, chemistry and even aspects of biology. Quantum theory is here to stay - at least until anyone finds anything belter.

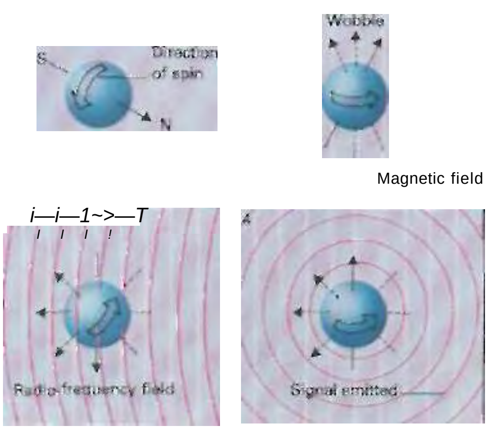
Nuclear magnetic resonance imaging





MA To produce an NMR image, such as this image of an adult head, the patient lies within the coils of a big superconducting magnet. Points within the body are selected by varying the strength of the magnetic field in three dimensions using coils above, below and along the magnet's axis. The radiofrequency field supplied by other coils make hydrogen protons resonate.

** -4 This electrostatic potential map of a water molecule is derived using quantum mechanics. It shows the strong negative potential below the oxygen atom, and positive potential around the hydrogens.*



T *Protons in hydrogen nuclei behave like tiny magnets, as a result of their spin 11). In a magnetic field, the proton magnets tend to align with the field, but continue to wobble Iprecess) about the field at a frequency that depends on the strength of the field 12). Pulsing the protons with a radiofrequency field that is oscillating at the precession frequency flips them into the opposite orientation (3). The protons soon flip back to their original orientations and radiate radio signals as they do so. The key to NMR imaging is that the frequency of these signals depends on location, because the field varies.*

Electron spin and wave mechanics

Planck quantized energy and Bohr quantized angular momentum, and each involved the use of an integer-or "quantum number" - to specify how many basic quanta of energy or angular momentum a system is allowed to possess. Wave mechanics gave meaning to these quantum numbers by showing that they are associated with solutions to the basic wave equation. Indeed, a solution describing a given state of electrons in an atom is associated with several quantum numbers. Thus each energy level corresponding to each line in an atomic spectrum can be defined in terms of a unique set of quantum numbers.

In 1925, two Dutch physicists, Samuel Coudsmit (1902-1978) and George Uhlenbeck (b. 1900), postulated another member for this set - a quantum number associated with an intrinsic angular momentum of the electron, in other words a "spin " quantum number. Goudsmit and Uhlenbeck were attempting to explain the fine structure of hydrogen spectra, in which single emission lines are seen on close inspection to consist of two closely-spaced lines. Their solution was to introduce the spin of the electron, which they asserted can exist in only two states, specified by quantum numbers + V2 and - V2. This is like saying that a spinning top can spin about the vertical in one of only two ways, either clockwise or anticlockwise. As a shorthand, the electron is said to have a spin of 'A. In 1928 the British theorist Paul Dirac (1902-1984) successfully united quantum mechanics with special relativity (4page 42) and showed that in the resulting theory a spin of 'A ***is*** *fundamental to the electron.*

The electron is not unique in possessing spin; the protons and neutrons in the nucleus also each have a spin of V2. This means that the nucleus itself can have a net spin angular momentum, depending on how the individual contributions from the protons and neutrons add together. The effect of the quantization of nuclear spin can be detected in atomic spectra as hyperfine structure, or splitting on scales a thousand times smaller than the fine structure splitting due to the electron's spin. (This is related to the large difference between the mass of an electron and the mass of a nucleus.)

*The fact that some nuclei, and in particular the proton, spin like children's tops has important applications. A proton (or any nucleus) has a positive charge, so a spinning proton is a moving electric charge and therefore generates a magnetic field (**i** page 57). In other words, protons and any nuclei with net spin are like subatomic magnets. This property is put to good use in the effect known as nuclear magnetic resonance (NMR). In **a** magnetic field, a proton wobbles about the direction of the field at a specific frequency. The art of NMR is to apply a field and measure this frequency. The technique is important to chemists because the precise frequency depends on the location of the proton - that is, a hydrogen atom - within a particular chemical. NMR is also valuable in medical imaging, because it shows the distribution of hydrogen across the body and can thereby reveal different tissues in a way that is potentially less hazardous than X-ray imaging.*

Near absolute zero, superfluid helium defies gravity to escape from a container

Quantum effects at low temperatures

At high temperatures, the average energy of the atoms and molecules is many multiples of frequency times Planck's constant, so the restrictions imposed by quantum theory on the vibrations of an atom about its lattice site in a solid pass unnoticed. At low temperatures the effect is dramatic. It is as if we could use only whole numbers of dollars to make purchases: the effect on the price of an automobile would be minor, but the cost of a newspaper would change drastically,

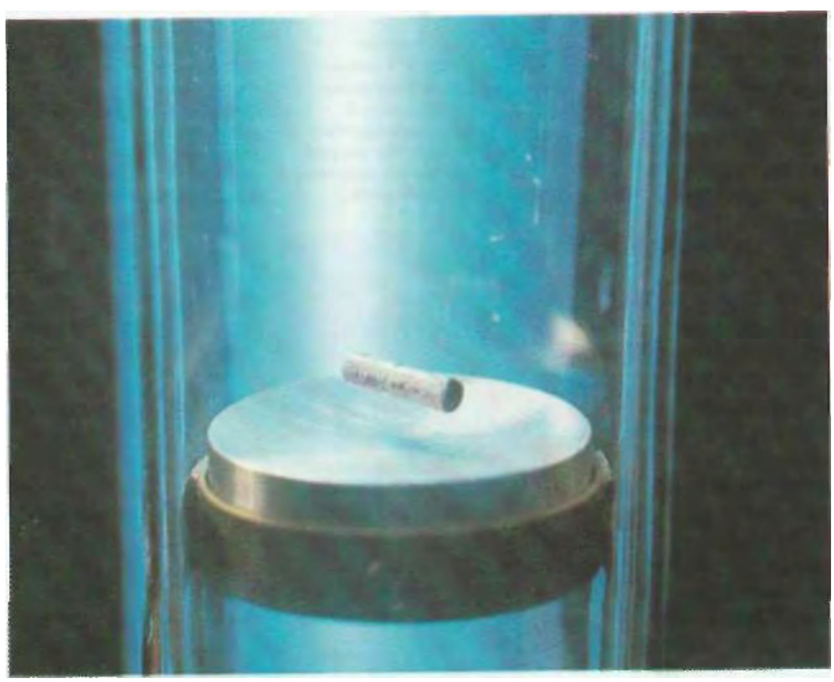
In 1907, Albert Einstein first applied the ideas of quantization to a theory of the heat capacity of solids. In 1912, the Dutch scientist Peter Debye (1884-1966) developed a simple theory that fitted well experimental data on the variation with temperature of a solid's heat capacity. At high temperatures, the spread in energy of the different atoms means that it is highly probable to find atoms with one or two units of energy more or less than the average value. When the temperature is increased a little, all that happens is that more atoms are to be found with higher energy than before. At very low temperatures the situation is different. When the temperature is raised, most of the atoms remain with the minimum allowable energy - their ground state -and only a few atoms are promoted to higher energy levels. Thus the solid can accept only a small fraction of the heat energy that it would accept at high temperatures.

Long before the introduction of quantization, many of the properties of metals were successfully described by the "free electron theory", in which the conduction electrons are not bound to a particular atom but are free to move about from atom to atom within the metal. However, the theory did not describe the magnetic and thermal properties very well. It was as if the electrons did not contribute to such processes. The major advance in understanding metals came only with the use of the "exclusion principle", due to the Austrian physicist, Wolfgang Pauli (1900-1958). The energies of electrons in a metal are also quantized, but unlike the atoms (or ions) which vibrate independently about fixed lattice sites, the electrons are all part of one system. According to Pauli's exclusion principle, no two electrons can have precisely the same energy. Once the lower energy states are filled up, the electrons must have higher energy to occupy the next available levels.

When the metal is heated, only those electrons in the highest energy level can be excited to the unoccupied higher energy levels; the electrons in the lower energy levels are unaffected because far too much energy would be needed to excite them. It is like an apartment block where the first family lives in the bottom apartment, the second family immediately above, family five on the fifth floor, and so on, with unoccupied appartments being at the very top of the building. If family two or seven wants to change apartment they cannot move one floor; they must go right to the top of the building. Only families already at the top of the building near to unoccupied levels can move easily. At room temperature only 1 percent of electrons contribute to the thermal properties of metals; no wonder that a problem existed before quantum theory.

>• The structure of a "high temperature" superconductor, VBp.:,Cu:O,,,:, displays its properties at temperatures that can be reached with liquid nitrogen.

T In the mid-1980s, research into the high-temperature superconductors found several such materials, offering great hopes for the applications of superconductivity.



• A bar magnet lowered onto superconducting tin creates a magnetic image in the dish, and is therefore repelled by the force between like poles, so that it remains levitated above the dish. The field of the magnet induces persistent currents in the tin, which form the magnetic image.

Superconductivity

Following the liquefaction of helium in 1908, the Dutch physicist Heike Kamerlingh Onnes (1853-1926) and other workers began a systematic investigation of the mechanical, electrical and thermal properties of many substances at temperatures a few degrees above absolute zero. Onnes discovered that at temperatures below about 4.25K, the electrical resistance of mercury abruptly disappears. Lead and tin also appear to lose their resistance or become "superconducting".

Onnes was able to show that the resistance falls to zero, not only to some very small amount. He set up a current in a coil of lead in a bath of liquid helium, and then isolated the coil from the battery. The current in the coil produced a magnetic field, which Onnes could monitor with a compass needle. The needle pointed steadily in the same direction, both before and after the battery was disconnected, indicating that the current encountered no resistance at all. In such circumstances, an electric current circulating in a loop will flow forever, so that no energy is required to maintain the current.

How does the resistance fall to zero? It is not possible to remove the obstacles that normally hinder the drift of electrons. Rather, in a superconductor the obstacles become ineffective; they no longer obstruct. In superconductivity, the drifting electrons become linked together in pairs via the lattice of atoms they travel through. These pairs are in the lowest possible energy state, and cannot lose energy when they meet obstacles. The interaction that creates these pairs is feeble: they are usually destroyed by an increase in temperature or by a high magnetic field. The material then becomes a normal conductor.

Until 1986, the highest temperature at which superconductivity had been observed was at23.3K, in an alloy of niobium and germanium - although for practical purposes alloys of niobium-titanium, which become superconducting around 9K, have proved most useful. Then, in 1986, two physicists working in Switzerland, Alex Mullerand Georg Bednorz, discovered a new class of materials - ceramic metal oxides - some of which are superconducting at much higher temperatures, around 100K. Such temperatures are reached with a refrigeration system that works with nitrogen (boiling point 77K) rather than helium (4.2K).

A Another spectacle from the strange world of low temperatures is the fountain effect - a result of one of the peculiar properties of superfluid helium. In its superfluid state, the liquid acts rather like a superconductor of heat, and flows toward a heat source as if attempting to eliminate it. To produce this effect, a heater in the glass vessel "attracts" superfluid helium which rushes into the vessel below the heater, and forces the jet of liquid through a small hole at the top.

See *also*
light 35-44
Atoms and Elements 65- 72
Studying tho Nucleus 79-86
Fundamental Forces 105-10

A The motion of a racing car at full speed is frozen by the camera, but at a price. In panning the camera to track the car, the background has become blurred. Similarly, but more fundamentally, at the quantum level the better motion is known, the more uncertain position becomes.

T The Austrian physicist Erwin Schrodinger formulated the theory of wave mechanics in the mid-1920s. He was greatly influenced by the work of the French prince, Louis de Broglie 11892-1987), who first proposed that particles such as electrons could be described as waves.

Heisenberg and uncertainty

In the early 1920s a German, Werner Heisenberg (1901-1976) was wary of taking the idea of electrons in fixed orbits too literally, and began to work on a theory that depended on observable quantities only. The result, published in 1925, was the first theory of quantum mechanics - a mechanics to describe the subatomic world. The theory was based on mathematical structures called matrices. It did not at first reach such wide recognition as Schrddinger's theory of quantum wave mechanics, published the following year. This was more easily understood, but in fact the two approaches are mathematically equivalent and lead to the same answers. Heisenberg was rewarded with the 1932 Nobel Prize for physics; the prize for the following year went to Schrodinger.

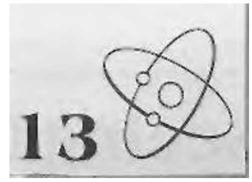
Two years after discovering quantum mechanics, Heisenberg went on to postulate his famous "uncertainty principle", in 1927. This states that certain pairs of physical quantities, in particular position/momentum and energy/time, can never both be known exactly. Indeed, the better one is measured, the more uncertain the other becomes. Imagine, for example, trying to measure the location of an electron with a powerful microscope that uses a beam of very short-wavelength gamma rays. According to quantum theory the beam consists of a stream of photons, which can reveal the electron by bouncing off it up the microscope. But each photon that bounces off the electron will change its momentum, just as in subatomic billiards. Indeed, the more accurately we try to pinpoint the electron, using gamma rays of shorter and shorter wavelength, the more momentum the individual photons give to the electron because their energy increases with decreasing wavelength.

Schrodinger and his cat

The Austrian physicist, Erwin Schrodinger (1887-1961) never liked the "dual" picture of matter that behaves both as waves and particles, and he tried to construct a theory based on waves only. He encapsulated some of his concerns about quantum theory in his "cat paradox". Schrodinger imagined a cat locked in a box with a phial of poison and a radioactive nucleus, and he supposed that the phial would be broken by the decay of the nucleus, in which case the cat would be killed. The question is, at any given time can anyone say what is the state of the cat- whether it is dead or alive - without opening the lid? The quantum physicist can only calculate the probability that the nucleus has decayed (4 page 82). Only when the lid is opened and an observation made, is the answer known.

So what is the state of the cat (representing a quantum system) between observations? Does it really exist in this interim period? And, to take an extreme view, does it exist only in the mind's eye of the physicist, even when being observed? Such questions tax the minds of philosophers and physicists alike. One view is that the present theory of quantum mechanics is not complete; that there are "hidden variables" of which physicists are as yet unaware but which describe the quantum system in a fully deterministic way. Another view is that all possible states between observations really do exist. In this "many universes" interpretation of quantum theory, the Universe splits each time there are different possible outcomes at a quantum level. Thus the real Universe consists of a host of separate universes, but each of us is conscious of only one at a time. Each time we observe we proceed at random to the universe corresponding to only one of the possible outcomes.

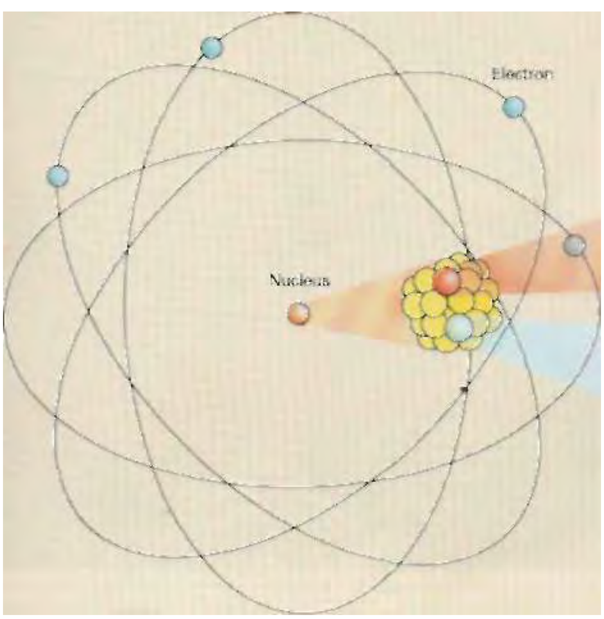
Elementary Particles



Building blocks of the nucleons...Quarks and antiquarks,..Leptons... Neutrinos.. .How elementary particles are made...PEFtsPECTivE...Naming the particles...Identifying particles...Particle accelerators and detectors...Antimatter...Cosmic rays...Predicting new types of particle

Since the beginning of the 20th century the picture of the fundamental nature of matter has changed dramatically. In the closing years of the previous century there were already signs that atoms could not be the indivisible, immutable objects that had once been imagined. In 1897 the British physicist Joseph John (JJ-) Thomson (1856-1940) discovered the electron, a tiny fragment of atomic matter about one two-thousandth the mass of an atom of hydrogen, the lightest element. Then with the discovery of radioactivity, physicists found that atoms could spit out other fragments - namely alpha particles (that is, helium nuclei) and so transform from one element into another. In 1911, Thomson's former student Ernest Rutherford (1871-1937) and his colleagues at Manchester University in Britain discovered that most of the mass of an atom is concentrated in a tiny central nucleus, which by the 1930s was known to consist of protons and neutrons.

Today the electron is still regarded as one of a small number of truly fundamental particles. As far as experiments can show, it behaves like a "point" of matter, with no internal structure, even down to distances as small as 10"^{19}m, one ten-thousandth the size of a proton. With protons and neutrons, however, the story is very different. Experiments in which energetic beams of electrons (and protons) are scattered by thin "targets" of various materials show that protons and neutrons are extended in space, having a definite diameter of some 10 "^{15}m. More detailed experiments, using electron beams of a high enough energy to probe within the protons and neutrons themselves, show that they contain an internal structure of pointlike objects. These "points" are particles called quarks and, like electrons, they appear to be truly fundamental.



Elementary particles

Proton

Quark

Neutron

Naming particles

The profusion of subatomic particles discovered since the late 1940s seems often to have been named with scant regard for logic. Yet beneath the bizarre terminology lies some sense. The first subatomic particle to be discovered - the electron - was given the name the British physicist George Stoney (1826-1911) had proposed several years previously for the "unit" of electric charge lost or gained by atoms during electrolysis. The proton is named for the Greek for "first", as it was the first component of the atomic nucleus to be identified. "Neutron", for the proton's neutral partner, then seemed a logical addition to the list.

With the discovery of a number of unpredicted particles by the early 1950s, Greek letters were introduced to label the different varieties. In the case of those initially labeled pi (n) and mu (/J.), the particles have become known as the pion and muon.

In 1962 two physicists, Murray Gell-Mann lb. 1929) and George Zweig (b. 1937) independently proposed that many of the growing number of subatomic particles, including the proton and neutron, must be composed of more fundamental objects. The protons and neutrons would each contain three of these objects, and Gell-Mann chose the name "quark", which occurs in the phrase, "Three quarks for Mister Mark", in the work "Finnegan's Wake" by James Joyce. Zweig chose the name "aces", but it was Gell-Mann's choice that stood the test of time.

Probing matter

Stream of electrons

_ Electron cloud

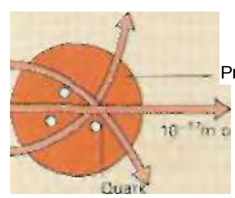
Q- ______Nucleus

10-"m

Deflection of electrons

100 MeV

10'GeV



Proton

*A Beams of electrons of increasing energy can be used to probe matter on smaller and smaller scales. The electrons, themselves electrically charged, are deviated, or scattered, by charged structures. At energies of 100 electronvolts (eV) or more, the electrons are scattered by the cloud of electrons around an atom. At energies a million times greater than this IWO MeV or more), the electrons penetrate this cloud but come under the influence of the positive nucleus at the heart of the atom. At energies of billions of electronvolts **(GeV),** the electrons scatter from the tiny charged quarks within the larger protons and neutrons.*

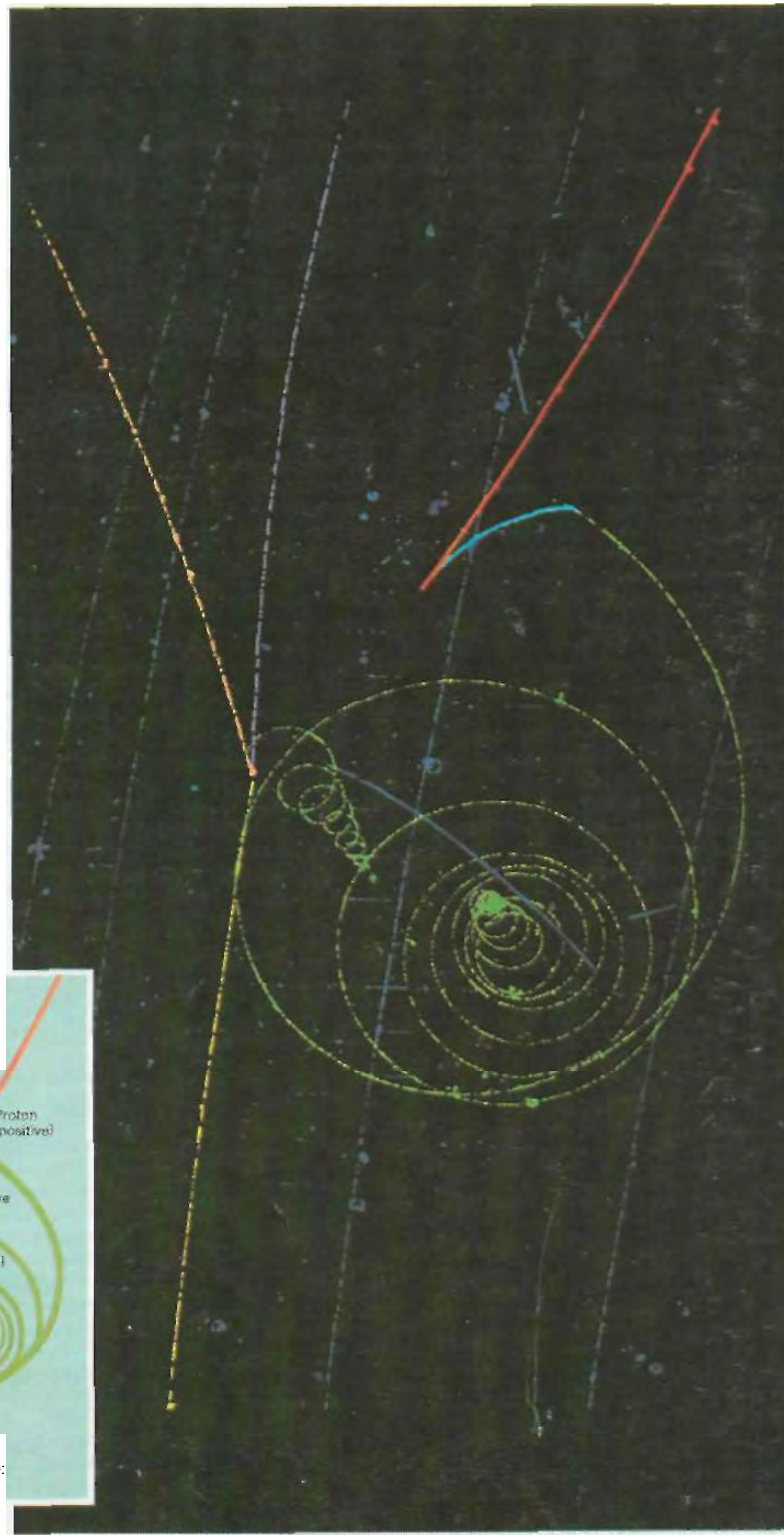
Seeing Inside the Atom

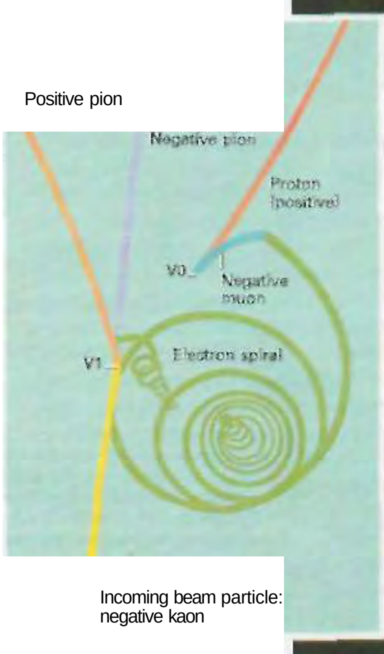
Accelerators

To see inside the atom requires techniques that can probe distances of 10^{-15}m and smaller. This is far beyond the capabilities of the most sophisticated microscopes, but the principles used are similar. An optical microscope uses light scattered from an object, but its resolution (or ability to reveal detail) is limited by the wavelength of visible light, around 10^{-7}m. An electron microscope forms images from energetic electrons scattered from an object. The higher the energy of the electrons, the shorter the wavelength associated with them, according to quantum theory.

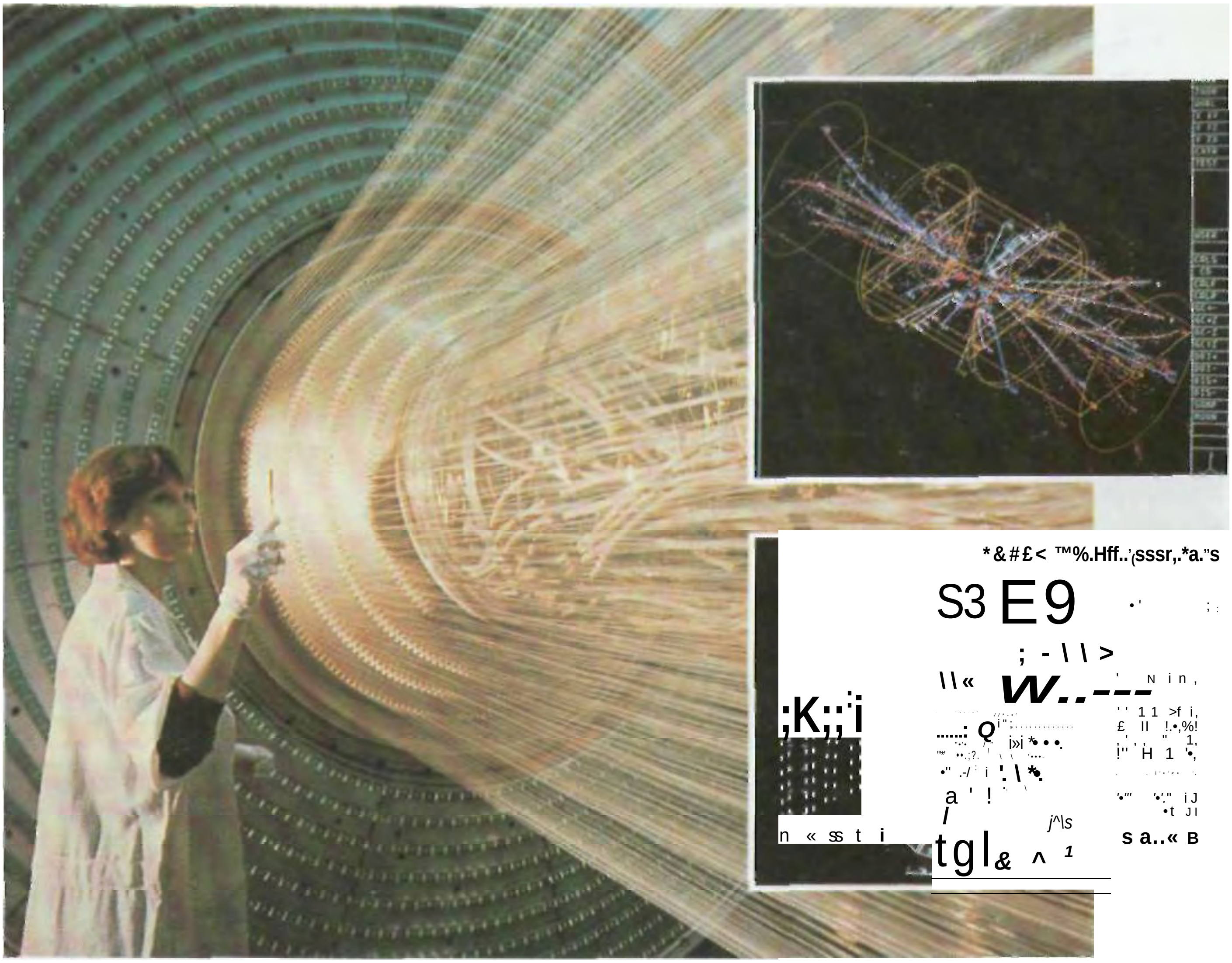
Studying matter on scales of 10^{-15}m or less requires very high energy beams of electrons, or other kinds of particle, produced at machines known as particle accelerators. Again, information about structure at these scales comes from observing the ways in which objects - in this case, subatomic particles within the nucleus - scatter the high energy beam. Although physicists do not reconstruct "images", the patterns of the scattered particles can be detected in specialized equipment. Moreover, the high energies can often reveal new, heavier lor more energetic) states of matter which cannot be observed at low energies.

Most particle accelerators have a format in which the beam travels in a circle through an evacuated pipe. Magnetic fields keep the particles on the correct path as they orbit the machine thousands of times. On each ciicuit, the particles pick up small amounts of energy from radio waves set up in sections of the machine known as the accelerating cavities. As the particles increase in energy, so do the magnetic fields guiding them until eventually the particles reach the maximum energy that the magnetic ring can contain. The beam is then directed out of the accelerator towards awaiting experiments, or to targets where it may be used to produce secondary beams of particles such as pions, muons or neutrinos.





• In this photograph of particle tracks in a bubble chamber colors have been added. Negative particles Ikaonsl cross the picture front below. Their tracks curve to the right showing that the chamber's magnetic field bends negative particles /purple, pale blue and green) to the right, positive ones lorange and red) to the left. The two spirals must be due to electrons, the only particles light enough to curl so much in the magnetic field. The tiny spiral is from an electron knocked from an atom in the liquid; the large spiral comes from the decay of the particle that made the pale blue track - this must have been a muon. The gap between the two "V-prongs shows that a neutral particle, which could not leave a track, has been produced at the first "V" and decayed at the second "V".



Bubble chambers

To work out what happens when a high energy particle interacts with a proton or a neutron within a nucleus, experimenters must record the tracks of the scattered particles. These often include many new particles created in the energy of the collision. Electrically-charged particles are relatively easy to detect. These leave trails of ionized atoms as they move through matter, and these trails can be revealed in a variety of ways. Moreover, charged particles are influenced by magnetic fields and the curvature of their paths through magnetic fields gives clues to their charge and momentum.

One device for tracking particles is the bubble chamber. This is a vessel containing a liquid under pressure, close to its boiling point. If the pressure is reduced, the boiling point should also drop. But if the pressure change is rapid enough, the liquid becomes superheated (remains liquid above its boiling point). If the pressure change occurs just after a beam of particles has passed through, however, the liquid is unstable enough to begin to boil along the path of the ionized particle tracks. Tiny bubbles form, revealing the paths of the charged particles, which can be photographed.

Wire chambers

Physicists using a bubble chamber must measure photographs of the tracks and feed the relevant numbers to a computer in order to calculate back to what happened in the original collision. Devices called wire chambers can eliminate one stage in this process by producing electrical signals that can be recorded directly in a form appropriate to a computer. In a wire chamber, charged particles leave ionized trails through a volume of gas, splitting the atoms into electrons and positive ions. Wires spread uniformly throughout the chamber are made electrically positive so that the negatively charged electrons are attracted towards the nearest wire. As the electrons approach the wire, they knock additional electrons out of the gas, until eventually an avalanche of electrons arrives at the wire, inducing a sizeable electrical signal. By piecing together the pattern of signals from a whole array of wires, a computer can reconstruct the paths of the original charged particles that passed through the wire chamber. If the chamber is within the coils of a large electromagnet the curvature of the tracks allows experimenters to measure the charge and momentum of the particles.

A *In wire chambers, thousands of wires cross the space traversed by the particles, as in this drift chamber (left} seen during construction. When complete, the wires in such a device are contained in a gas-filled vessel. Charged particles ionize the gas, releasing electrons which travel toward the wires under the influence of an electric field, and produce signals in circuits attached to the wires. Computers then reconstruct the track of each particle and display them in images. The top picture shows a three-dimensional view of tracks in a cylindrical drift chamber. Below is a cross-sectional display from a different experiment; this not only shows tracks in a drift chamber (red lines) but also "hfts" in other types of detector which help in identifying the different particles.*

Antimatter is created along with matter in equal amounts, but on meeting again matter and antimatter annihilate

The neutrons and protons in stable nuclei can live forever (or at least, for an extremely long time; * page 108). However, the energetic collisions of protons and nuclei reveal other subatomic particles. These are similar in many ways to the proton and neutron, but they are short-lived and cannot form part of the atoms and matter of the familiar world. These particles are also built from quarks, and analysis of the relationship between them has shown that there are probably at least six types of quark in all. The six are called "up", "down", "charm", "strange", "top" and "bottom" - although evidence for the top quark is still scant.

Particles such as the neutron and proton are built from three quarks, and are known collectively as "barvons". The "up" quarks and "down" quarks are the lightest varieties, and so together they form the lightest known baryons - the proton and neutron. The other four kinds of quark - strange, charm, bottom and top - are successively heavier and they form heavier particles. The neutral particle called the lambda contains an up quark, a down quark and a strange quark. It is like a neutron but with a heavier strange quark replacing one of the down quarks. This makes the lambda about seven percent heavier than the neutron. It is also unstable. After a life of only 2×10^{10} seconds it decays, usually to a proton and a particle called the pion, or to a neutron and a pion. In each case, the strange quark transmutes into a lighter quark and the difference in mass is taken away by the pion that has been produced at the same time. This transmutation of quarks is the same basic process that underlies the decay of a neutron to a proton in beta-radioactivity (| page 83). In beta decays, a down quark in the neutron transmutes into an up quark, thus forming a proton. Such changes between quarks bring all the heavier baryons down to the level of the proton, the lightest baryon of all.

There are also six antiquarks which form antibaryons such as the antiproton and antineutron. The lightest antiquarks are readily created in the collisions of protons and nuclei, in cosmic rays as well as in particle accelerators. However, the creation of antibaryons is quite costly in terms of energy, so the antiquarks pair up more easily with quarks created at the same time out oT the available energy. The resulting particles, built from a quark and an antiquark, are called mesons, meaning "between", since the first known examples had masses between that of an electron and that of a proton.

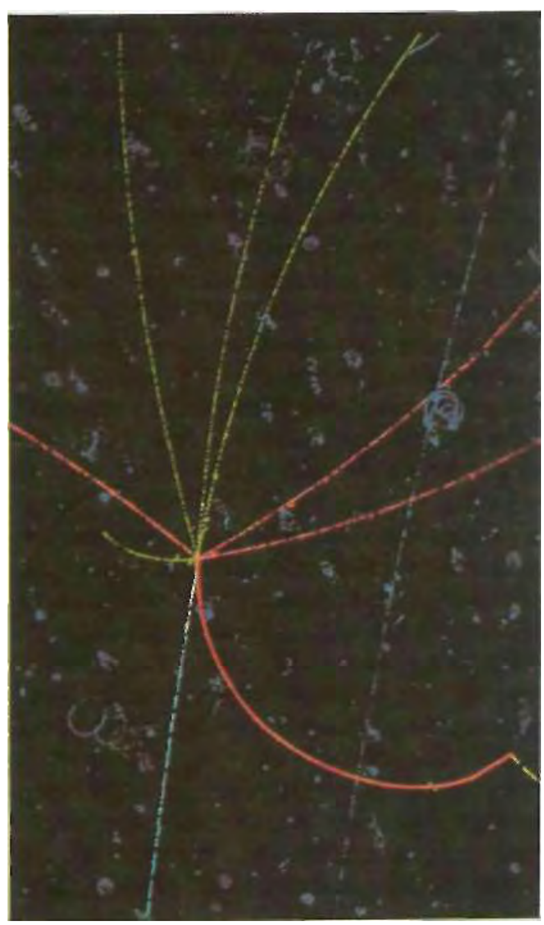
In charged mesons the quark and antiquark are of different types. For example, the positively charged version of the pi-meson, or pion, is built from an up quark and a down antiquark. Its antiparticle, which has negative charge, is built from a down quark and an up antiquark. Neutral mesons, on the other hand, can be built from a quark and the antiquark of the same variety. In this case the quark and antiquark temporarily orbit around each other rather like an electron and a proton in a hydrogen atom. However, such neutral mesons are shorter-lived than their charged counterparts. The neutral pion is a mixture (at the quantum level) of up-antiup and down-antidown. The mesons are all short-lived and ultimately decay to lighter particles. The quarks arc held together within baryons and mesons by the strong nuclear force - the strongest of nature's four fundamental forces (• page 106). It is due to the nature of the strong force that the quarks can form only clusters of three quarks (baryons) and quark-antiquark pairs (mesons). Other combinations are forbidden, including the possibility of single quarks. In practice, therefore, the quarks and antiquarks appear always to be confined within baryons and mesons.

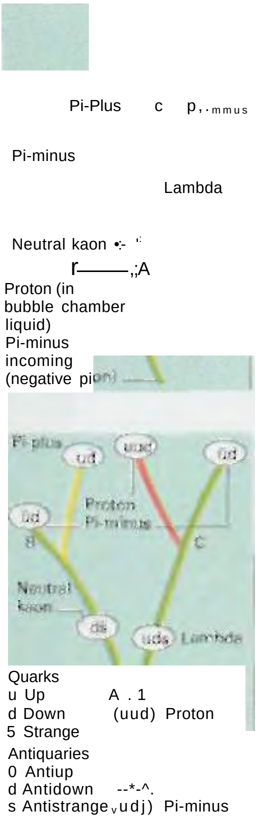
Antimatter

One of the main discoveries of modern physics concerns the existence of antiparticles. These are particles that have the same mass as the familiar electron, proton and neutron and the other more exotic species, yet which have exactly opposite properties such as electric charge. Thus the antielectron, or positron, has positive charge, while the antiproton has negative electric charge.

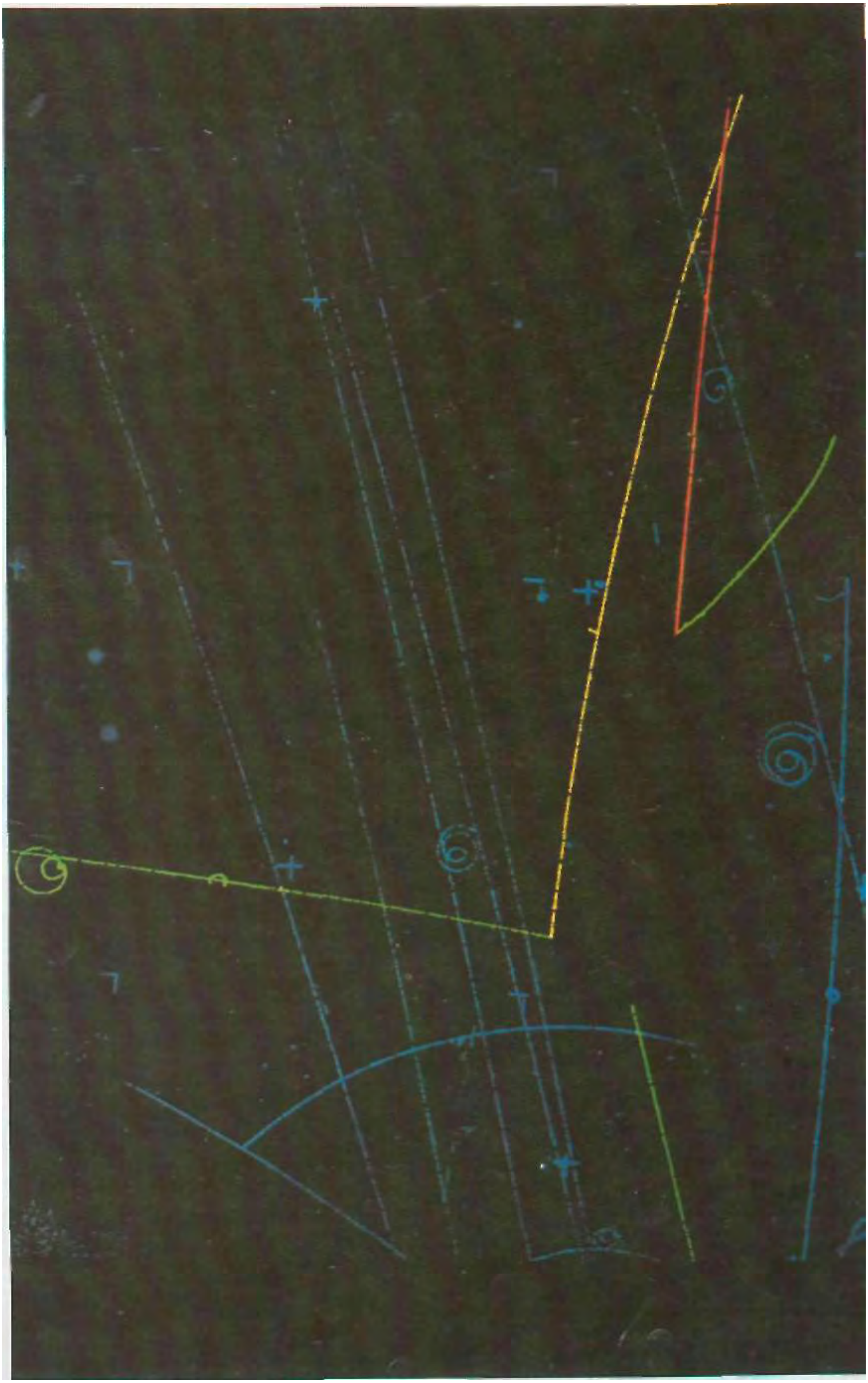
Antiparticles were postulated in a theory due to the British physicist Paul Dirac (1902-1984). In 1928 he combined quantum theory (jpage 87) with the special theory of relativity (4 page 42), but found that his equation describing the electron had two possible solutions. One described the familiar electron, while the other turned out to describe the antielectron, or positron. This interpretation did not become clear until 1932 when physicists studying cosmic radiation discovered tracks due to positively charged electronlike particles in detectors called cloud chambers. These particles occurred in "showers" with equal numbers of electrons. The conclusion was that electron-positron pairs were materializing from gamma rays generated as the cosmic rays passed through sheets of metal.

Dirac's theory also revealed the fate suffered by electrons and positrons that come too close to each other. They mutually self-destruct in a reversal of the process that creates electron-positron pairs, releasing an amount of energy equivalent to their total mass. This process is known as annihilation.





4 •- *A bubble chamber photograph shows the "associated production" of strange quarks and antiquarks. The long tracks are due to pions crossing the chamber; one, colored green, interacts with a proton and produces two neutral particles which leave no tracks (A). One contains a strange quark, the other a strange antiquark. The overall "strangeness" is thus zero both before and after the reaction. This is a property of reactions that occur via the strong nuclear force. The two particles betray themselves when they decay (B, C) via the weak nuclear force, and in this case strangeness can change. In the decay of the lambda particle, for example, the heavier strange quark gives birth to an up quark, a down quark and an up antiquark (C).*



M *An antiproton leaves a track in a bubble chamber (pale blue) before eventually annihilating with a proton. Their properties exactly cancel, creating a burst of pure energy. This rematerializes in the form of entirely different particles and antiparticles: four positive pions (red) and four negative pions (green).*

Elementary particles

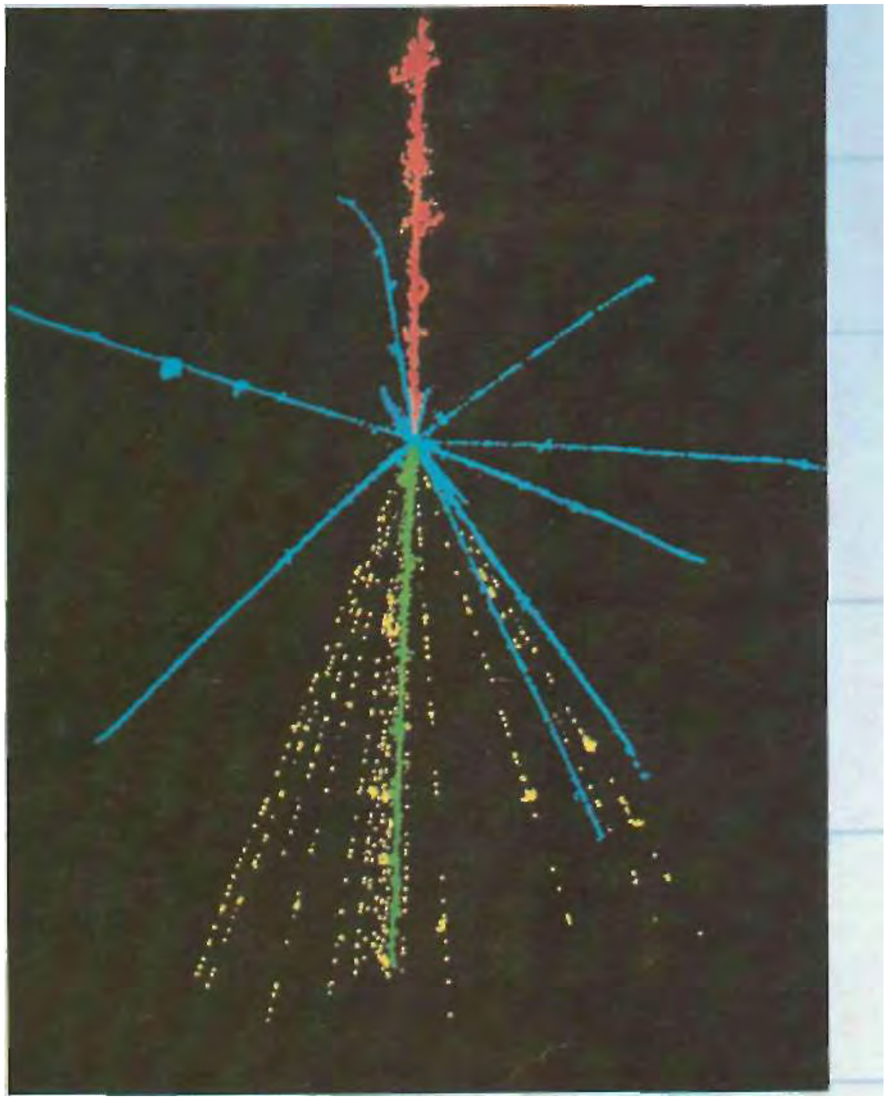
Baryons		Quarks	Mass (MeV)	Lifetime (s)
Proton	P	uud	938.3	$>10^{32}$y
.Neutron	n	udd	939.6	898
Lambda	A	uds	1115.6	2.6x1<r10
Sigma		uus	1189.4	0.8x10-'°
	V	uds	1192.5	5.8X10-20
		dds	1197.3	1.5x10 '°
Xi	=0	uss	1314.9	2,9x10"10
(cascade)	H	dss	1321.3	1.6x10"'°
	A	udc	2282.0	2.3x10-'3

Quarks		Charge	Mass (MeV)
Up	u	+•%	5
Down	d	-V5>	7
Charm	c	+%	1400
Strange	s	-'A	150
Top	I	+%	
Bottom	b		1,800

Mesons				
Pion	-n1-	ud.ud	139.6	2.6X10"8
		(uu.dd)	135.0	0.8X10"'6
Kaon	K*	us.su	493.7	1.2X10-8
	KS	(ds.sd)	497.7	0.9X10-10
	K»			5.2X10-8
	D*	cd.cd	1869.4	9.2x10-'3
	D°	cu	1864.7	4.4X10-'3
	F-	cs.cs	1971	1.9X10-'3
	B*	ub.Ob	5270.8	14X10"'3
	B°	bd	5274.2	

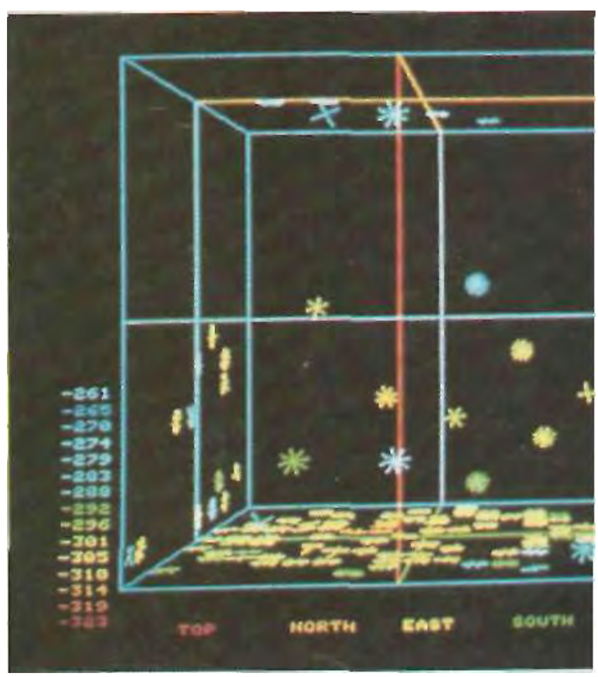
d, u.s.c, 6,-Antiquarks

Luplunu		Charge	Mass (MeV)	Litetimes (s)
Electron	e	-1		stable
Electron-neutrino	VQ	0	<0.46x10"4	stable
Muon	u.	-1	105.6	2.197X10-6
Muon-neulrino	v1^	0	<5.0	stable
Tau		-1	1.784	3.4X10-'3
Tau-neulrino	vT		<164	stable



X.

The quarks are only one type of fundamental particle. The electron is clearly not built from quarks. Instead it belongs to a second family of fundamental particles - the leptons. This name comes from the Greek for slight, for at one time all leptons seemed to be lightweight. Like the family of quarks, the family of leptons contains six members. There arc three negatively-charged members: the electron, the muon and the tau. And there are three neutral members which are called neutrinos, and which seem to have little or no mass. However, neutrinos are clearly of distinct types. In particle interactions one sort will produce only electrons, one sort will produce only muons, and the third variety should give rise only to taus, although this has still to be proved. They are referred to as the electron-neutrino, the muon-neuirino and the lau-neutrino. As with the quarks, the six leptons all have corresponding antiparticles, which are positively charged in the case of the antielectron (positron), antimuon and antitau.



The electron and the three types of neutrino seem to be stable, but the muon and tau can both decay to lighter particles. The muon, which has only about 11 percent the mass of the proton, can decay only to an electron, because no other charged particles lighter than the muon exist. When the muon decays, il also gives rise to a muon-neutrino and an electron-antineutrino. The tau's mass is nearly double

Primary cosmic ray

Nuclear interaction::.'.

<10,000m

Charged pion

Nucleon

25,000m

Neutral pion

Gamma rays

20,000n

Muon

5,000m

Neutrino

Muon

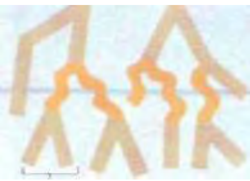
Neutrino

Electron-positron pair

2,000n

Ground level

• Tracks of cosmic rays have been captured in photographic emulsion and colored later. A sulfur nucleus (red) had collided with a nucleus in the emulsion to produce a spray of particles including a fluorine nucleus (green), 16 pions and several other nuclear fragments (blue).*



Electron-positron pair

< Showers of secondary cosmic rays are produced in the atmosphere when high energy particles and gamma rays from outer space (the primary cosmic rays) collide with atomic nuclei there. The primary particles reflect the general composition of the galaxy, and are mainly hydrogen nuclei (protons) but also, more rarely, heavier nuclei. The secondaries include a variety of subatomic particles, such as pions, which are very short lived and decay to muons and neutrinos. The muons often survive to reach ground level and can penetrate below the surface of the Earth. The neutrinos are weakly interacting and can even pass though the Earth.

Cosmic rays

A rain of subatomic particles, together with electromagnetic radiation that includes ultrahigh-energy gamma rays, arrives at the top of the Earth's atmosphere from space. It comes from the Sun, from stars, and possibly from beyond our Galaxy. The high-energy part is the cosmic radiation, or cosmic rays. The particles are mainly protons, but there are also nuclei typical of the matter abundant in the Solar System, from helium to uranium.

The cosmic rays collide with nuclei high in the atmosphere, and can produce great showers of subatomic particles at ground level. It was in attempting to discover the nature of the cosmic radiation that physicists in the 1930s and 1940s first came across new varieties of subatomic particle which are not stable constituents of the atoms of everyday matter. The American physicist Carl Anderson (b. 1905) discovered the positron - the first example of antimatter-in 1932, and later took part in recognizing the existence of the particle now called the muon. In 1947 the British physicist Cecil Powell (1903-1969) and his colleagues found the first evidence for the pion.

These discoveries encouraged physicists in the 1950s to build particle accelerators, which could mimic the effects of cosmic rays but in better controlled conditions and at far greater intensities. Experiments at such machines led to a multitude of mainly short-lived particles. This work culminated with the idea that most particles observed in such experiments are not fundamental, but are built from a number of elementary particles - the quarks.

•< This display records the collision of a cosmic ray neutrino in a detector consisting of 10,000 tonnes of ultrapure water. The tank, in a mine 600m below ground, has walls lined with light-sensitive phototubes. This view shows the phototubes that fired with the light from particles created when the neutrino collided in the water.

that of the proton, so it has access to many different decay routes. It can even decay to particles containing quarks, provided that they together contain equal numbers of quarks and antiquarks.

There seems to be a natural symmetry between the two families of fundamental particles. Indeed, the resemblance runs even deeper because each family can be divided into three "generations". The matter of the everyday world is built from up quarks and down quarks (in atomic nuclei) and electrons. In addition, electron-type neutrinos and antineutrinos are "emitted in the radioactive decays of some unstable nuclei. Thus only two quarks, one charged lepton and one neutral lepton are required to explain the everyday world.

However, studies of cosmic rays and at accelerators reveal muons and their neutrinos, and strange and charmed quarks. Once again this group of four particles contains two quarks, a charged lepton and a neutral lepton. Indeed, this "second generation" of quarks and leptons seems to mirror the first almost precisely,:, but in a higher-energy world. The quarks and the charged lepton of the second generation are all heavier than their counterparts in everyday matter. Lastly, since the mid-1970s, experiments at higher energies Have suggested a third generation of still heavier quarks and leptons. This should contain the bottom and top quarks, the tau and its neutrino.

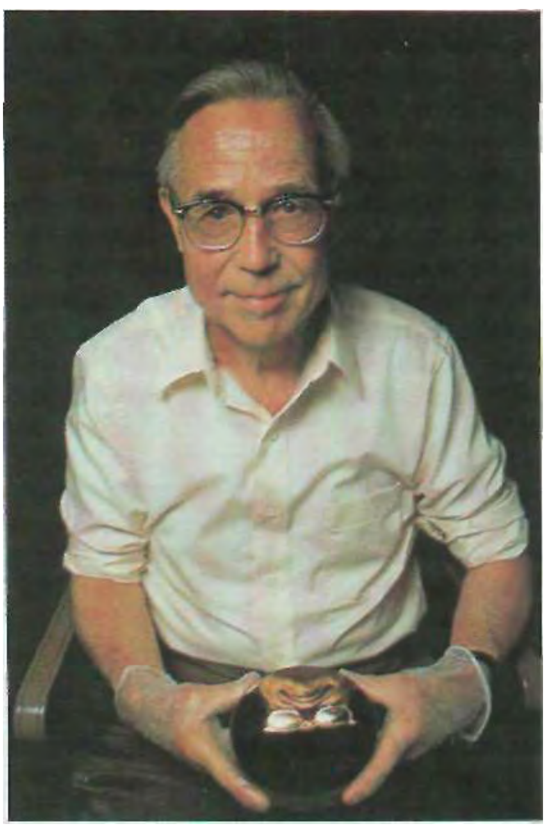
See also
Forces, Energy and Motion 11-20
Studying the Nucleus 79-86
The Quantum World 87-96
Fundamental Forces 105-10
Nuclear Fission and Fusion 119-24

free quarks and monopoles

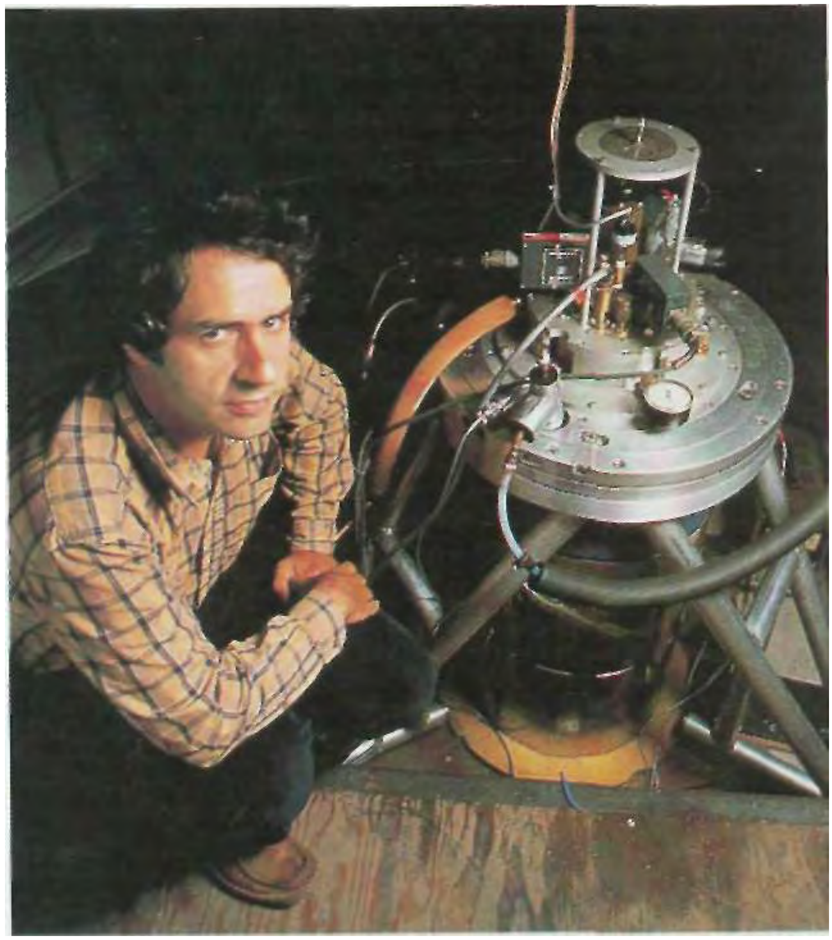
The quarks have electric charges that are fractions of the size of the electron's charge. The up quark has a positive charge of% the size of the electron's (negative) charge, while the down quark has a negative charge of V3 the electron's charge. If single quarks exist, this "fractional" charge should reveal them as other particles have zero or integer charge.

There have been no convincing examples of particles with fractional charge. One intriguing result came from United States physicist William Fairbank (b. 1917), who found evidence for charges of V3 located on tiny balls of niobium. He levitated the balls magnetically, and then made them oscillate in an alternating electric field. By measuring the amount of oscillation he found that on some occasions some balls carry a charge of V3. Many physicists remain skeptical of these results.

In 1982 Bias Cabrera (b. 1946) claimed evidence for the monopole. Monopoles are particles that carry a single magnetic charge or pole. Though predicted by Dirac's theory of 1931, there is no experimental evidence that they exist. Theory suggests that monopoles were formed in great numbers during the Big Bang. Some of these "relics" may still be around. As time passes and new experiments fail to find any new evidence, the existence of these particles remains in doubt.



f William Fairbank at Stanford University has observed indications of fractional charge when he levitates tiny spheres of niobium between two plates like the one he holds.



»• Bias Cabrera, in his basement laboratory at Stanford University, together with an improved, three-coil version of his monopole detector. He has recorded one event in 1982.

Antimatter such as positrons occurs naturally in the showers of cosmic rays that cascade through the Earth's atmosphere. Antiprotons and antineutrinos can also be created. But the everyday world is manifestly built from electrons, protons and neutrons, and it seems that the nearby region of the Universe must all be built from matter rather than antimatter, or there would be evidence of large releases of energy every time pieces of matter and antimatter encountered each other. However, in the collisions of cosmic rays and in experiments, particles and antiparticles are always created in equal numbers. So why is there no evidence of large amounts of antimatter in the Universe?

Some theories of the forces at work in the very early Universe suggest that matter and antimatter were indeed created in equal amounts in the hot Big Bang with which the Universe began. However, subtle asymmetries in the behavior of the fundamental forces between particles (| page 105) tipped the balance in favor of matter. While most of the matter and antimatter annihilated, a small excess of matter remained to form the Universe that exists today.

Why should matter have replicated the fundamental particles at least three times, revealing only one set in the physical reactions of life on Earth? The answer to this basic question may be lied up in an understanding of the evolution of the Universe. In the Big Bang, the heavier generations would have existed on equal terms with the familiar quarks and leptons of the first generation. The total number of generations may have had an influence on how the Universe developed. One of the challenges to particle physicists is to search for evidence of additional generations of the quarks and leptons.

Fundamental Forces

14

The forces of the Universe: Strong nuclear force, weak nuclear force, electromagnetic force and gravitational force...Bosons, particles carrying the forces... PERSPECTtvE...Einstein's theory of general relativity ...Gravitational lenses...Grand Unified Theories and the Theory of Everything...Proton decay experiments

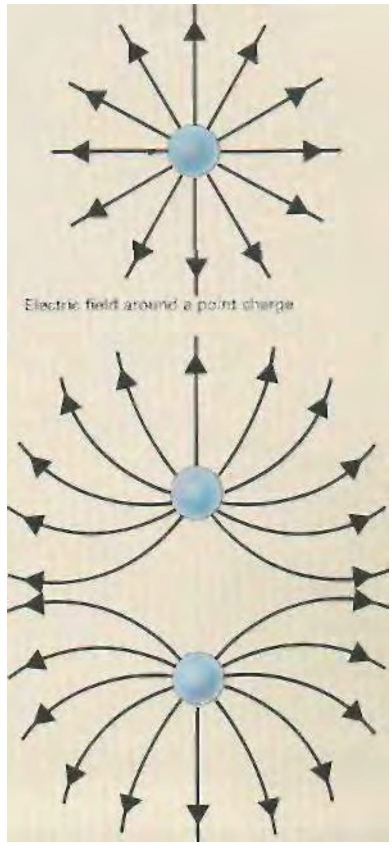
One surprising feature of the physical world is that much of its great diversity can be described in terms of only two fundamental forces - the gravitational force and the electromagnetic force. Gravity controls the motions of the stars and planets, and will even determine the fate of the Universe as it expands from its initial Big Bang. The electromagnetic force keeps electrons tied to the nuclei of atoms, and governs the motion of electrons in all kinds of systems, ranging from the "silicon chip" to the human nervous system. However, there are two additional fundamental forces, which are less apparent because they operate only within the confines of the atomic nucleus. These are the weak nuclear force and the strong nuclear force.

Modern theories of these four forces are all generally based on the mathematical framework known as a "gauge field theory". Each force is regarded in terms of a field describing the strength and direction of the force throughout space and time. (The term "gauge" is related to the concept that measurements at different locations in space and time should always give the same results). An important feature of these theories is that there should exist particles which in a sense "carry" the force. These are called gauge bosons. The term "boson" refers to the fact that these particles have integer values of the intrinsic angular momentum, or spin (| page 93). This makes the gauge particles fundamentally different from the particles of matter - the quarks and leptons - all of which have a spin value of *Vi*.

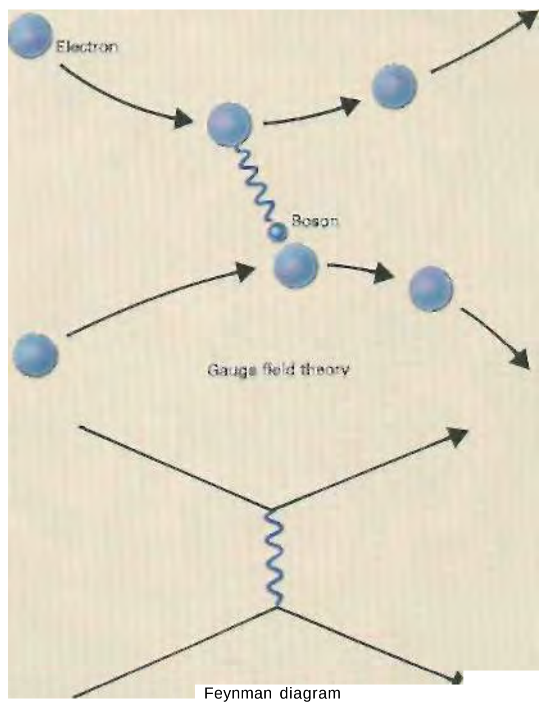
The best developed gauge theory, called quantum electrodynamics or QED, deals with the electromagnetic force. It accounts for the behavior of charged particles even down to the level of the fundamental quarks and leptons (| page 98). In QED the source of the electromagnetic force is electric charge; its gauge bosons are the familiar photons of light. Charged particles interact by exchanging photons between themselves, rather as in a game of quantum football.

Force	Particles affected	Range	Relative strength	Bosons (particles exchanged)	Role in Universe
Strong	Quarks	10-'Bm		Gluons	Holds quarks together withir protons and neutrons and other baryons and mesons
Electromagnetic	particles	Infinity :	10'''	Photons	Determines structures of atoms, molecules, solids and liquids
Weak	Quarks and^ leptons	10-^m	10~8	W, *Z*, particles	Determines stability of atomic nuclei: fuels Sun and stars
Gravitational	All	Infinity j	Itr"0		Assembles matter into planets, stars and galaxies

Quantum electrodynamics



Electric field between two similarly charged points

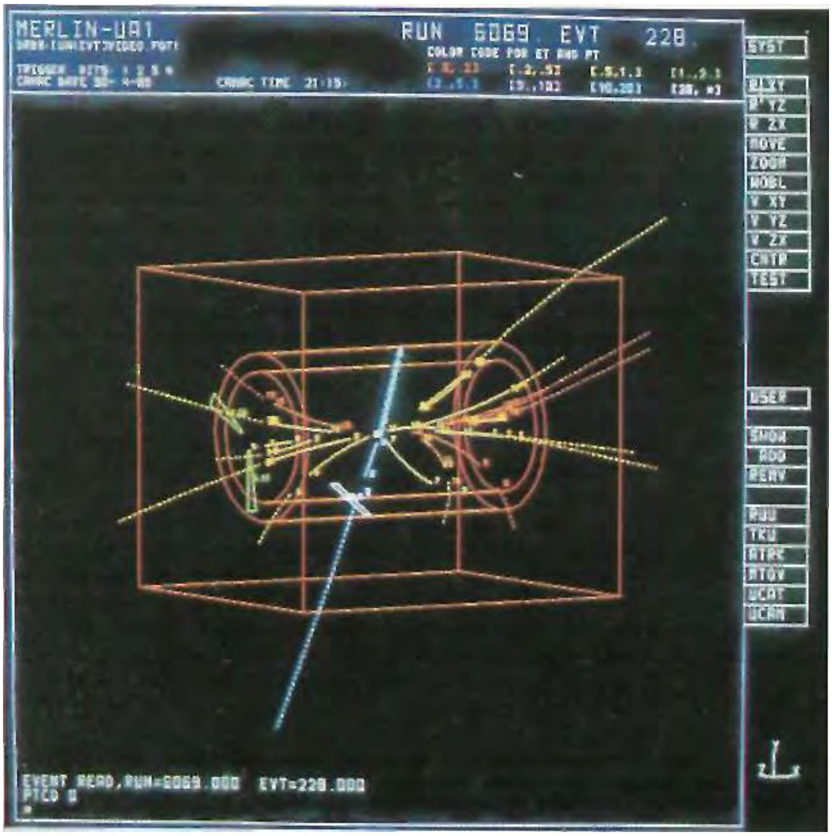


Feynman diagram

M Physicists recognize four basic forces that appear to underlie all phenomena, from atoms to the workings of the Universe. Each has its own exchange particle.

A At the quantum level, the electric field between charged objects is mediated by an exchange particle - the photon -and depicted in Feynman diagrams.

The pion was predicted to exist in theory 12 years before it was actually detected



Pions and the strong force

The first viable theory of the strong nuclear force was put forward in 1935 by the Japanese theorist Hideki Yukawa (1907-1981). The electromagnetic force was already understood in terms of the exchange of photons at the quantum level and Yukawa considered a similar mechanism for the strong force. He argued that, as the strong force acts only over very small distances, it must be carried by a heavy particle. (The electromagnetic force, with infinite range, is by contrast carried by a particle of zero mass - the photon). Yukawa proposed that the carrier of the strong force must be a particle with a wavelength equal to about 10~^{5}m, and that the mass of such a particle is about 15 percent of the mass of the proton. He proposed that such particles are responsible for binding protons and neutrons together as they are constantly exchanged.

Yukawa's particle - now known as the pion - was discovered in 1947. His picture of pion exchange is still useful; however, the strong force is now known to originate in the quarks within the protons and neutrons. The binding of the protons and neutrons in a nucleus is due to a leakage of the strong interquark force.

•4 A proton and an antiproton have collided at the center of an electronic detector, producing some low energy particles (yellow and red tracks! and a W particle. The Whas decayed to an electron (long blue track) and a neutrino (arrowed).

A The lightweight particles called pions are produced in abundance in energetic nuclear collisons. Here a high energy proton (yellow) collides with one at rest in the liquid in a bubble chamber. The collision produces as many as seven negative pions (blue) and nine positive particles /red) which include seven positive pions.

4 This photographically colored image shows the decay of a Z particle - the neutral carrier of the weak nuclear force and partner of the W particle. Again, a proton and an antiproton have collided at the center of the image, producing a melee of tracks. Two of the tracks (yellow) are particularly straight. They belong to an energetic electron and positron from the Z's decay.

The weak and strong nuclear forces

During the 1960s QED was incorporated into a larger theory in an attempt to describe the electromagnetic force and the weak nuclear force at one and the same time. The successful outcome is now known as the electroweak theory. The weak nuclear force is responsible for the decay of particles such as the neutron and hence underlies beta-radioactivity (4 page 82) and was recognized as distinct from the strong force that binds the nucleus in the early 1930s. It is carried by three gauge bosons - a neutral particle called Z, a positive particle called W^+ and a negative particle called W".

Unlike the photon, which has zero mass, these particles are all heavy, being 90-100 times as massive as the proton. It is the large mass of these particles that makes the weak force weak. In exchanging a gauge boson, matter particles (quarks and leptons) must temporarily violate the principle of energy conservation (| page 18). This can occur at the quantum level through the uncertainty principle (4 page 96), but it is allowed for only a short time - so short that the weak force acts only across distances the size of the nucleus or less, and with a strength some 100,000 times weaker than the electromagnetic force within ordinary matter.

"Colored" particles

The strong force acts within the same domain as the weak force, but it is very different in its behavior. It is the strongest of nature's fundamental forces, and within the nucleus it dominates the electromagnetic force by a factor of more than 100. The gauge theory of the strong force is called quantum chromodynamics, or QCD. According to this theory, the source of the strong force is a property called "color" (this has, however, nothing to do with the color of light), and to carry the strong force there are eight gauge bosons called "gluons", which act between "colored" particles. Quarks carry the property of color, but leptons do not. This explains why baryons and mesons (which contain quarks) interact strongly with each other, while leptons (which do not contain quarks), do not interact through the strong force with baryons, for example.

Color can be thought of as similar to electric charge, but it differs in that it occurs in three different varieties - called red, blue and green after the primary colors of light - where electric charge occurs only as positive and negative. An important rule of QCD is that quarks can form only those clusters that have a total "color" of zero - or white in the analogy with real colors. Thus only clusters of three quarks (baryons) and quark-antiquark pairs (mesons) exist in the world at large. A baryon, for instance, must contain a red quark, a blue quark and a green quark; it cannot, say, contain two red quarks and a green quark. Moreover, single quarks, which are by definition colored objects, cannot exist alone - and indeed there is no convincing evidence for single quarks, despite many searches (| page 104).

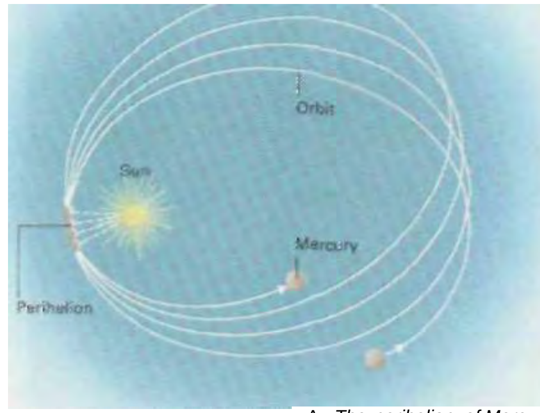
The rule about color is tied in with the nature of the gluons, the gauge bosons of QCD. A gluon, like a photon, has no mass, but they differ in a crucial way. The photon has no electric charge: it is not a source of the electromagnetic force that it conveys. Gluons, on the other hand, are colored, and as a result they can interact among themselves via the strong force. This has die profound effect of making the strong force stronger at greater distances within a particle, and also of confining die colored quarks within the dimensions of the subatomic particles that they form.

Theoretical physicists can provide a quantum theory of gravity-if they posit a Universe often dimensions

Gravity and the general theory of relativity

Newton's theory of gravity can be used to calculate motions above the Earth, from the path of a ball to the orbit of a satellite. It also reveals how the gravity of the Moon and the Sun produces tides on Earth. However, despite its many successes, Newton's law of gravitation has failed in some small but crucial areas.

Mercury is the closest planet to the Sun. Its orbit is an ellipse, and the closest point of approach to the Sun on this orbit is called the "perihelion". In successive orbits, this point advances slightly around the orbit. The rale of advance is very slow - about 0.159° per century. The gravitational forces from the other planets cause about 0.147° of this advance, but Newton's laws cannot explain the remaining 0.012°.



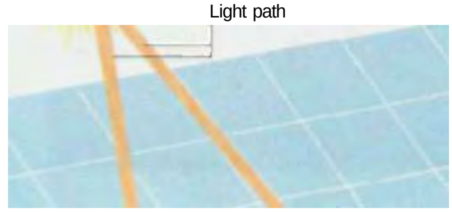
A *The perihelion of Mercury - the point of the orbit closest to the Sun - moves slightly from one orbit to the next. Newtonian gravity can account for most of this effect, but to explain it completely required the application of Einstein's general theory of relativity.*

The explanation for the discrepancy did not come until more than 200 years after Newton, with the work of the German physicist Albert Einstein (1879-1955). Einstein is famed for his special theory of relativity, one of the basic assumptions of which is that no information can be transmitted faster than the speed of light *(4* page 42). According to Newton, however, gravitational forces act instantaneously throughout all space. The two theories disagree.

These problems stimulated Einstein to look again at the theory of gravity. In 1907 the Russian mathematician Hermann Minkowski (1864-1909) combined the three dimensions of space with the dimension of time, and proposed that all events in the Universe occur in a four-dimensional "space-time".

Quasar

Light path



Einstein adopted Minkowski's idea of space-time, and proposed that space-time is curved in the presence of massive bodies. All particles, including light, travel on the shortest route (a geodesic), and where space-time is curved these routes are curved. According to Einstein, mass produces a curved space-time which makes particles move as if attracted. This has been summarized as, "Matter tells space how to curve and space tells matter how to move".

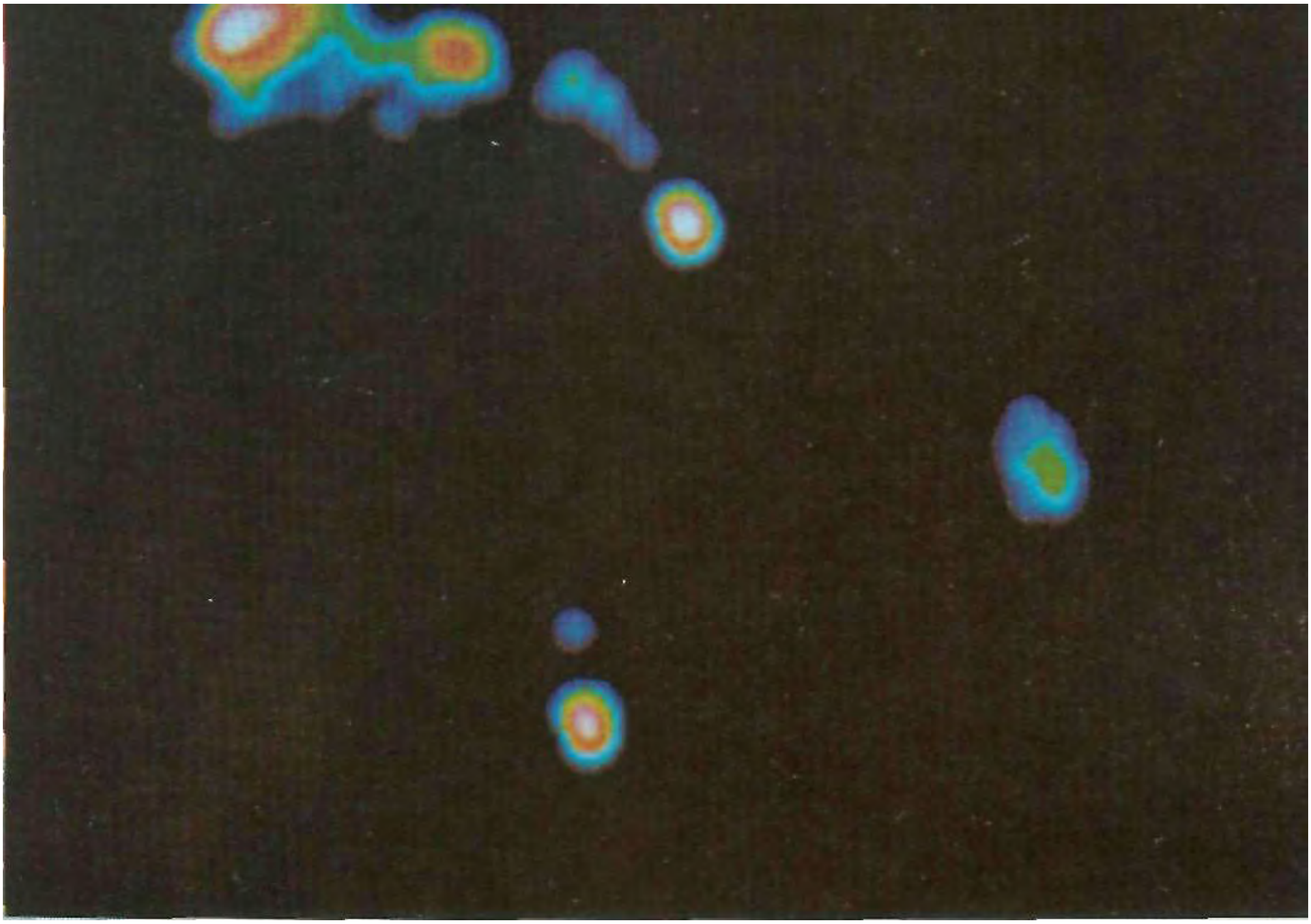
Einstein's theory is known as the general theory of relativity. The German theorist Karl Schwarzchild (1873-1916) solved Einstein's equations to discover the shape of space-time in the presence of a massive spherical object. His solution can be used to reveal the planetary orbits around the Sun. For most practical purposes Einstein's orbits are identical to those of Newton, but Einstein's theory predicts the extra 0.012° advance in the perihelion of Mercury, which Newton's theory was unable to explain.

Gravitational lens

....._ Galaxy

Another prediction of Einstein's theory proved a crucial test. Light always travels by the quickest path. This means that, in the curved space-time near the Sun, light should be deflected by 0.0049°. Newtonian gravitation predicts a deflection of half this. A way to measure the deflection is to observe a star whose light passes close to the Sun. This is possible only when there is a total solar eclipse. In 1919, the British astronomer Arthur Eddington (1882-1944) organized an expedition to Principe, an island in the Gulf of Guinea, and to Brazil, to measure the deflection of light during the total eclipse of the Sun that year. The results, first announced at the Royal Society in London, were in complete support of Einstein's theory.

Three hundred years after Newton, gravity is still not properly understood. Physicists in many countries continue to work on gravity to understand what causes it, and how it is transmitted. Some physicists are trying to merge gravity with the quantum theories of the subatomic world, in an attempt to show that all Nature's forces are only different manifestations of a single underlying force.



* A *A large mass warps the space around it so that light passing the object does not travel in a straight line, but follows a bent path rather as when traveling through a lens. The effect of such a "gravitational lens" is believed to have been observed in the "double quasar", 0957+561. In the radio telescope image above, the two bright white blobs correspond to objects that both appear to be quasars at identical distances, 10,000 million light years away; it seems likely therefore that they are two images of the same object. The upper image is thought to be the direct view of the quasar, while the lower is presumed due to light bent around an intervening galaxy.*

Earth

Superstrings

The dream of many theorists is to find a "theory of everything " (TOE) that incorporates gravity along with the strong and electroweak forces. But so far gravity has defied a complete description at the quantum level. There is no equivalent ofQED for gravity. It is possible, however, to use general relativity to hypothesize the properties of a gauge boson for gravity. This is the graviton, and it differs from the photons, gluons, and W and Zparticles in that it has spin 2; the others have spin 1. However, while there is good evidence for the other gauge bosons, there is none at all for the graviton.

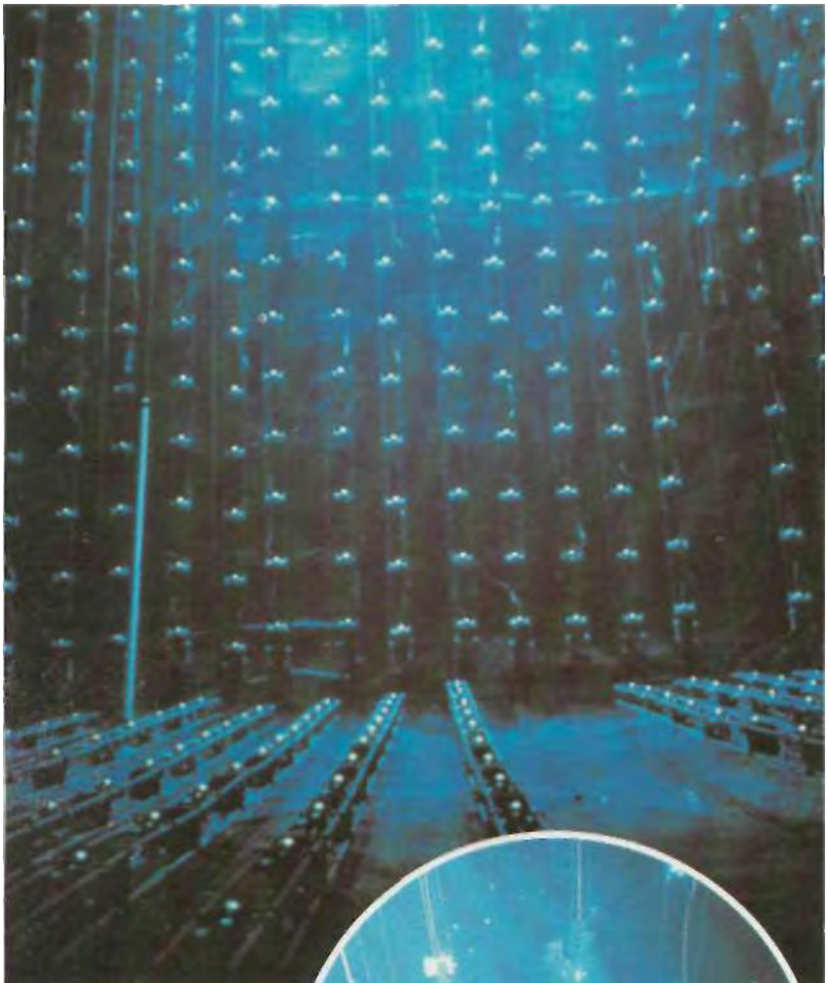
Most attempts to build a quantum theory of gravity have proved inconsistent in a fundamental way, but in 1984 a British theorist, Michael Green, and his United States colleague John Schwarz, discovered a class of theories that appear to avoid these inconsistencies. The theories are types of "superstring " theory, in which the fundamental particles - or quanta -are represented as tiny string-like objects. This contrasts with the "point-like" representations of conventional quantum field theories. The strings are about 10^{i3S}*m long, a dimension that is far smaller than can be studied with the most powerful particle accelerators. But on the larger scale at which physicists can make observations, the strings appear like points, just as in the conventional theories.*

Superstring theories contain a symmetry called supersymmetry-hence the name superstrings. Supersymmetry directly relates bosons (force-carrying particles with spins 0,1,2,...) to fermions ("matter" particles such as quarks and leptons, with spins '/?, %,...). This symmetry has an elegance that is attractive to theorists because it unites these apparently disparate sets of particles together; it also seems to be a crucial ingredient of superstring theories.

Superstring theories appear at first to have the drawback that they work only in 10 dimensions - that is, they describe a universe with six dimensions over and above the familiar three dimensions of space and one of time. But this proves not to be a problem, for the idea of extra dimensions is not new. Theorists know how to make the unwanted dimensions disappear by "curling up " so that they are not apparent in the everyday world.

The remarkable discovery of Green and Schwarz in 1984 was that two - and only two - specific types of superstring theory appear to be properly consistent. The inclusion of gravity in these theories does not have the devastating effect it has in other theories. Whether superstrings prove to be the key to the ultimate theory of everything remains to be seen, but they offer one of the most exciting advances in theoretical physics for many years.

See also
Light 35-44
Etectromagnetism 57-64
Studying the Nucleus 79-86
The Quantum World 87-96
Elementary Particles 97-104



A > One test of the GUT is that it should be possible to *detect the occasional decay of protons. An experiment to do this involves a "cave" 600m below Lake Erie, Ohio, filled with water and lined by phototubes. These tubes would register the particles produced if a proton decayed.*

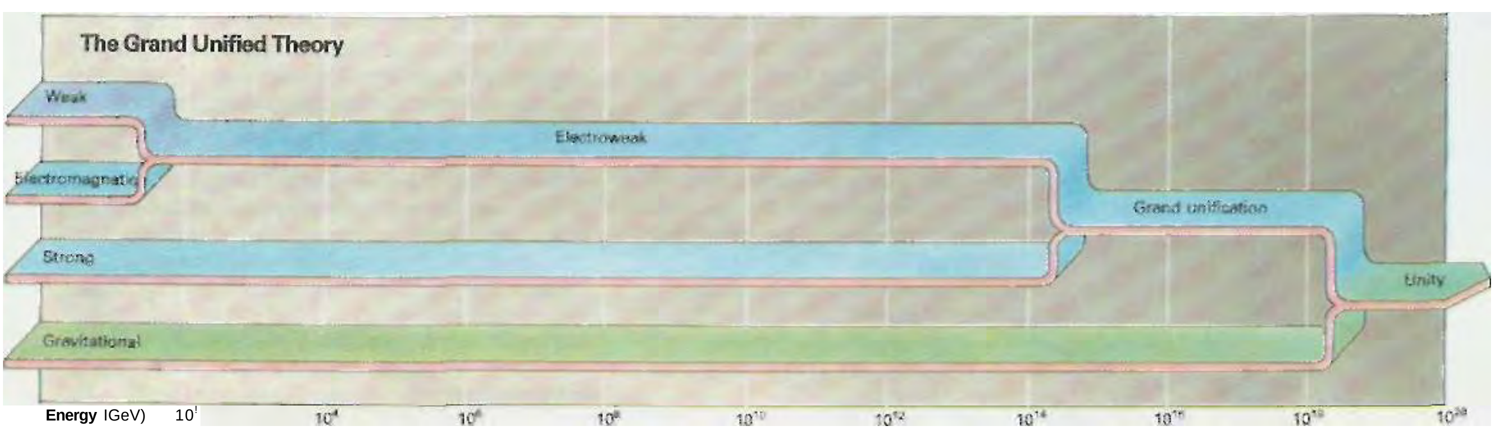
J The four fundamental forces eventually combine to a single force at ever higher energy levels.

Grand unified theories

The electroweak theory "unifies" two of nature's four fundamental forces. It has inspired theorists to incorporate it with the gauge theory of the strong force - QCD - to form a single "grand unified theory" (GUT) that would embrace the strong, weak and electromagnetic forces. Such theories generally link quarks and leptons in a way that implies that quarks can decay into leptons, albeit very rarely in the everyday world, at which energies are generally low. One important implication is that protons should not be stable, but have an average lifetime of some 10^{32} years or more. As this is an average it should be possible to monitor a large enough collection of protons and observe a few decays in the space of a year or so. Such experiments, involving hundreds of tonnes of water or some other material, are underway in many parts of the world. However, none of them has yet found any convincing evidence for the predicted decays of protons.

Another important feature of unified theories is that they contain an inherent symmetry between the forces they unite. Electroweak theory predicts that, at energies higher than those typical in atomic nuclei, the weak force becomes stronger until it eventually equals the electromagnetic force in strength. At these higher energies the WandZ particles can be exchanged as readily as photons, and the two forces appear as symmetric facets of a single electroweak force. These ideas have been verified in high-energy experiments at particle accelerators, where Wand Zparticles have been produced and observed. In a similar way, grand unified theories predict that at a very high energy - far beyond the reach of present accelerators - the strong and electroweak forces become symmetric.

Ideally, theorists would like to include gravity in a single "theory of everything " (TOE). In such a theory, all four forces would be symmetric facets of a single underlying force. Only in the extreme heat of the Big Bang at the start of the Universe would energies have been high enough for the four forces to appear the same. As the Universe expanded and cooled, each of the forces would have separated out, first gravity, then the strong force, and finally- at energies now reached in particle accelerators— the weak and electromagnetic forces parted company. The original symmetry of the Universe would have become hidden in the diversity apparent in its cool, low energy state of today.



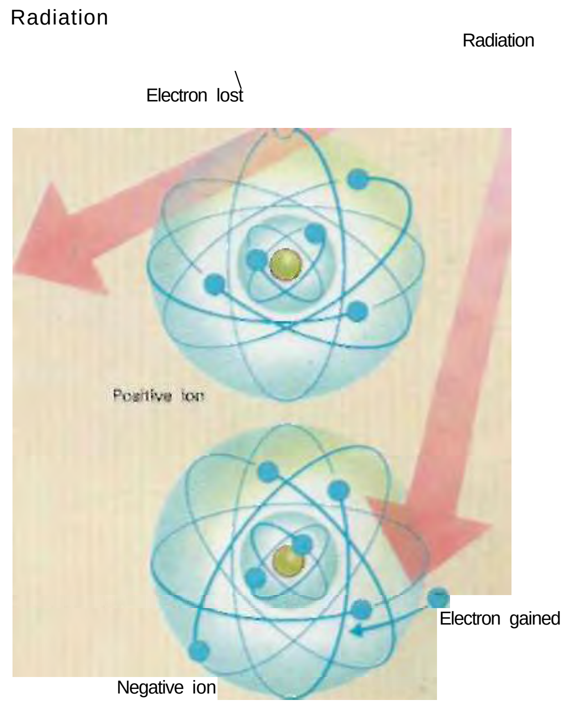
Radiation and Radioactivity

15

Alpha, beta and gamma radiation...Stopping radiation...Natural sources of radiation—Effects of radiation on the body...Using radiation to see inside the body...PERSPECTIVE...Pierre and Marie Curie... The discovery of X-rays...Units of radiation and dosage... Predicting the effects of low-level exposure

Radiation is all around. It rains down from outer space, and leaks out of the ground and from the walls of buildings. It is almost totally natural. Only 13 percent of the radiation dose thai people receive in Britain, for example, comes from artificial sources, and most of thai is in medical uses. Much less comes from the burning of coal, the testing of nuclear weapons, and only 0-1 percent of the total comes from radioactive wastes and effluents. But what is radiation?

The term radiation can refer to electromagnetic waves (from gamma rays to radio waves: 4 pages 60-61) or to energetic subatomic particles. However, in common use and especially in the context of radiation protection, the term refers more specifically to ionizing radiations - in other words, radiations that lose energy by ionizing atoms (knocking electrons out of them) or sometimes even by disrupting atomic nuclei. Such radiation can be electromagnetic or it can consist of particles. The particle component includes the alpha rays (helium nuclei: two protons and two neutrons) and beta rays (electrons) emitted in radioactive decays (4 page 83), as well as the showers of cosmic-ray particles generated in the atmosphere by "primary" cosmic rays from outer space. Strictly speaking only charged particles can ionize, because the basic reaction is due to the electrical force between the charged particle and the atomic electrons. But some neutral particles, in particular neutrons, and the X-rays and gamma rays at the high-energy end of the electromagnetic spectrum, are indirectly ionizing, because they produce charged particles when they interact with matter.

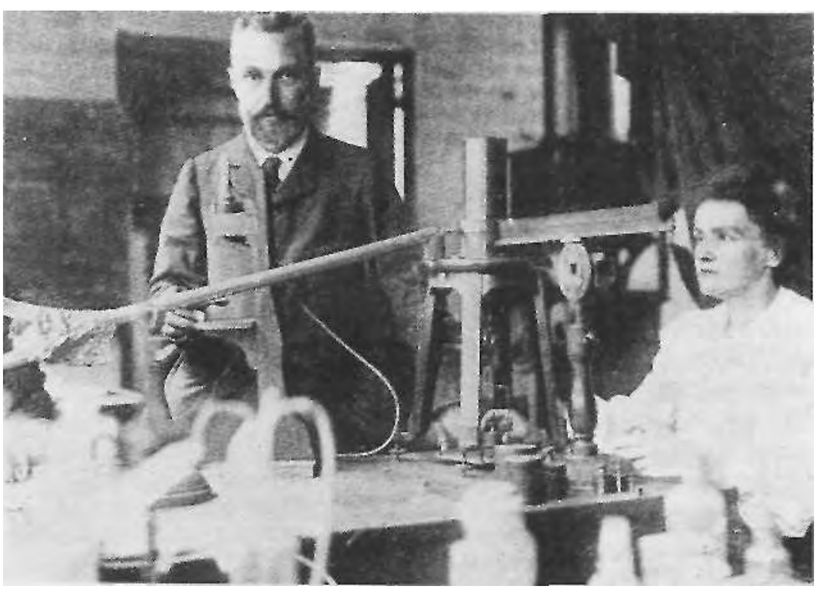


• Radiation causes damage when it knocks electrons out of atoms (left), creating positive ions. The electrons can then attach to other neutral atoms and create negative ions.

The Curie family

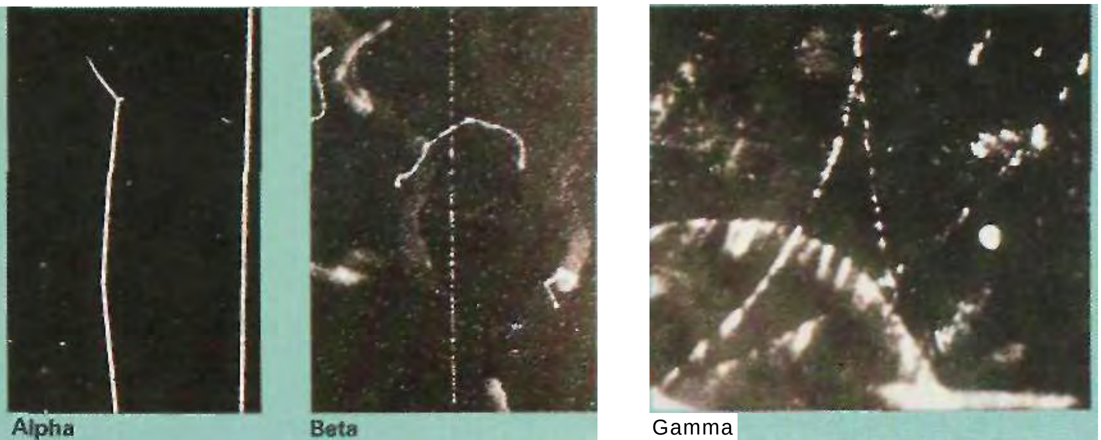
The French physicist Henri Becquerel (1852-1908) discovered radioactivity in 1896, when he found that radiation from a uranium salt would darken a photographic plate kept well away from light. He was soon able to show that uranium was the element responsible for the emissions. However, it fell to two other scientists working in Paris, Marie Curie (1867-1934) and her husband Pierre Curie (1859-1906), to show that other elements could produce similar rays. In 1898 Marie found that thorium emitted the same kind of radiation, and she also discovered that uranium ore (pitchblende) emitted more radiation than pure uranium. Together the Curies laboriously purified kilogram after kilogram of pitchblende, and in 1899 they announced the discovery of two new elements, which they called polonium, after Marie's native Poland, and radium. They also gave the effect the name still in use today- radioactivity.

Pierre Curie was tragically killed in a road accident in 1906, but Marie continued to do much important work on radioactivity, identifying beta-radiation as a stream of negatively charged particles. In 1903, Marie and Pierre shared the Nobel Prize for physics with Henri Becquerel for their work on radioactivity. Marie later became one of the few people to receive two Nobel Prizes, when she was awarded the prize for chemistry in 1911, for her discovery of radium and polonium.



• Pierre and Marie Curie in their laboratory in 1898 during their early work on radium. The instrument here is a quartz electrometer, which they used to measure the amount of ionization in air caused by various radioactive samples.*

• *Alpha particles are heavily ionizing - their tracks in a cloud chamber (a type of particle detector) are thick and short. Beta rays (center), which consist of energetic electrons, are poorly ionizing and leave thin tracks. Gamma rays (right) are non-ionizing and leave no tracks, but give rise to pairs of electrons and positrons I f page 100), which are ionizing.*



A • *Rontgen's discovery of X-rays in 1895 captured the world's imagination. Their ability to "see " inside things was appreciated as a valuable medical aid, but the harmful effects of even quite low-level exposure was less clear.*

Ionizing radiation

The effects of ionizing radiation on materials vary greatly with the type of radiation. The higher the charge of a particle, the more rapidly it ionizes. Alpha particles with the same velocity as protons lose energy four times as rapidly as the protons because they have double the charge of the protons. And the lower a charged particle's velocity, the greater its rate of ionization. As a charged particle loses energy it slows down and the energy it loses through ionization increases, until suddenly the particle no longer has enough energy to ionize an atom. The particle becomes neutralized and comes to a halt; it is said to have come to the end of its range.

Alpha particles from radium-226, for example, have enough energy to travel about 20 micrometers in aluminum. Beta particles (electrons), on the other hand, are far more penetrating. This is because for the same energy an electron, which has only 0-01 percent the mass of an alpha particle, has a much higher *velocity,* and energy losses are inversely proportional to the square of the velocity. An electron from the beta decay of carbon-14 travels about 100 micrometers through aluminum. If it had the same energy as the alphas from the radium-226 it would travel as far as one centimeter.

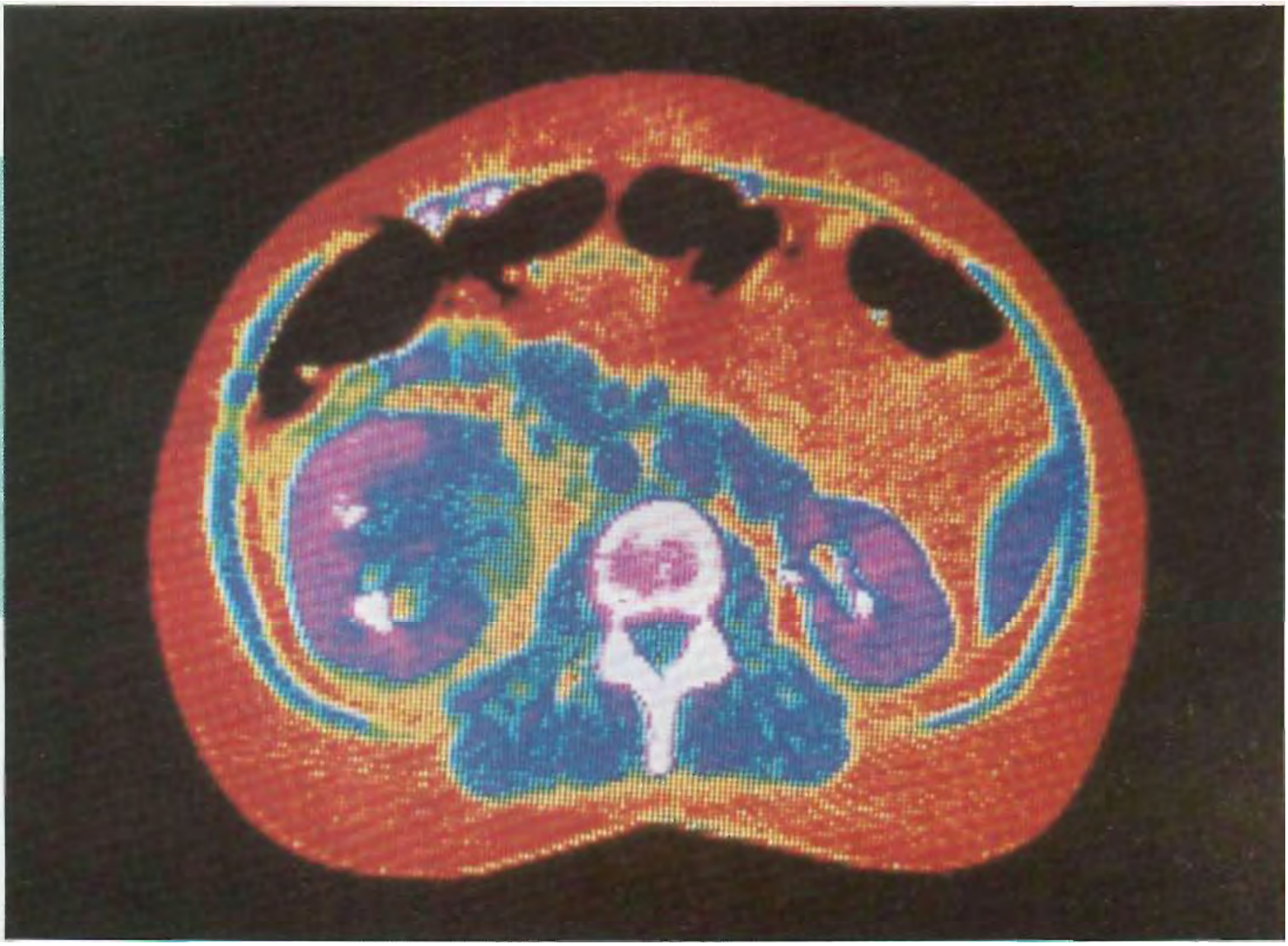
X-rays and gamma rays can also penetrate materials. Indeed, X-rays

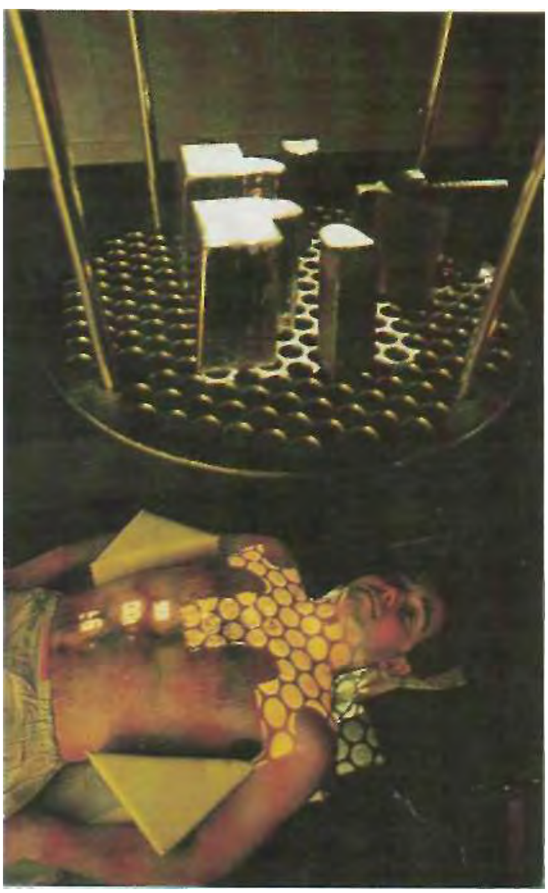
Rontgen and X-rays

On 8 November 1895, Wilhelm Rontgen (1845-1923), professor of physics at the University of Wurzburg, was studying the passage of electricity through a gas at low pressure in a cathode-ray tube. This was a glass tube with a positive electrode (a metal wire) at one end and a negative electrode at the other. As the gas was pumped out of the tube, the residual gas would glow, and eventually the wall of the tube opposite the cathode (the negative electrode) would glow green. Scientists had been investigating these phenomena for several years, but on this particular evening Rontgen noticed for the first time that a nearby screen was fluorescing, even though the tube was covered in black card. Invisible rays emanating from the cathode-ray tube were causing salts of barium platinocyanide to fluoresce. Over the next few weeks Rontgen discovered that the rays would ionize the air, would darken photographic plates, and, most remarkably, would pass through a variety of materials that are opaque to light. He announced his discovery ofX-rays, as he called them, at the beginning of 1896, and immediately the rays captured the imaginations of scientists and the public alike.

The X-rays in Rontgen's experiment originated at the place where the "cathode-rays" within the tube struck the glass wall opposite the cathode, making it glow green. Two years later, in 1897, the British physicist J.J. Thomson (1856-1940) showed the cathode rays to be streams of electrons, emitted from the cathode (4 page 68). What Rontgen was observing was the emission of high-energy electromagnetic radiation — X-rays — from atomic electrons in the glass wall, which had been raised to higher energy levels by the cathode rays and were now returning to their usual energies.

Rontgen received the first Nobel Prize for physics in 1901. By this time X-rays were well known, and tubes were made commercially to produce the rays for medical use, where they were especially useful for observing fractured bones and metal objects lodged inside the body. Today, medical X-ray imaging can be a highly sophisticated technique. X-rays are also used in radiotherapy at far greater intensities to damage tissue deliberately in the treatment of cancerous growths.





< X-rays damage body tissues by knocking energetic electrons from atoms. This effect is put to good use in directing X-rays to kill tumor cells. Here a patient lies beneath lead blocks that define the area that X-rays will reach.

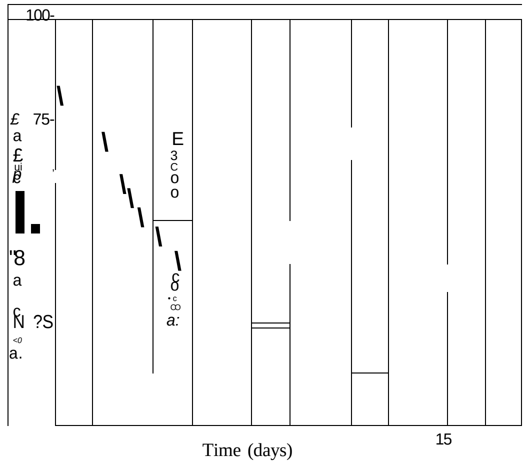
A Simple X-ray images, like the one shown on the opposite page, record the net effect of absorption by all the organs and tissues between the source of X-rays and the photographic plate. The more refined technique of computer assisted tomograph y (CA T) allows images to be formed of "slices" through the body, as in this example showing a section through the trunk -the spinal cord appears as white near the bottom of the image. CAT scanners work by rotating the X-ray tube around the body to define the "slice ".

are well known for this ability. The energies of X-rays are generally typical of the energies between electron shells in atoms (| page 71), and an X-ray is absorbed when its energy is sufficient to eject an electron from its shell. This electron may receive enough to leave the atom entirely and to ionize other atoms in the vicinity. In this way, the X-ray has the same general ionizing effect as a charged particle. The probability that an X-ray is absorbed varies approximately with the fourth power of a material's atomic number. Thus, lightweight elements, such as make up skin and muscle, transmit X-rays more easily than the heavier elements, as are found in bone or metals.

Gamma rays are of too high an energy to be absorbed easily by atomic electrons. They do, however, lose energy in collisions with the electrons. The discover}' of this effect by the United States physicist Arthur Compton (1892-1962) played a part in establishing the particle-like nature of electromagnetic radiation (4 page 88). These collisions can give the electrons enough energy to ionize. The collisions also reduce the gamma ray's energy, so that eventually die gamma ray becomes absorbed like an X-ray, but only after penetrating much farther than an X-ray can. Gamma rays from radioactive decays can pass through 10 centimeters of aluminum, a far greater distance than the alpha particles or even the electrons from radioactivity.

Most of the radioactive substances contained in the early Earth have long since decayed away completely

The uranium and thorium that occur in the ground and in the materials used for buildings provide the greatest natural exposure to ionizing radiations, along with potassium-40. This isotope has a half-life of 1-3 billion years, but forms only OT percent of all potassium. However, it emits gamma rays of relatively high energy, and it is the gamma radiation from radioactive materials that has the greatest effect on humans. Alpha particles, for example, cannot penetrate the dead outer layers of human skin whereas gamma rays can penetrate much farther. The uranium and thorium decay chains and potassium-40 are also responsible for most of the radioactivity taken internally in food and drink. Within the body, of course, alpha particles and beta particles produce greater effects. Here the alpha particles from polonium-214, one of the members of the uranium-238 decay chain, have a dominant effect, along with the gammas of potassium-40. However, the biggest effect of building materials lies in the



• *This body, well-preserved in the bogs of central England, has been radiocarbon dated to about* AD *200.*

T *Radiocarbon dating is used to determine the age of an organic substance by measuring the proportion of C-14 in a sample. The sample is cleaned and then oxidized in a sealed chamber, so that the C-14 is converted to radioactive* CO_2*. This gas is then placed in a counting chamber, comprising a mass spectrometer 14 page 67), for about 20 hours. During this time a sample 5,000 years old- the oldest that can be dated by this method- might register a count of 22,000 C-14 atoms.*

•* Radioactive isotopes are characterized by their half life -the time for half the nuclei in a sample to decay. The graph here shows how a sample of the isotope radon-222 decreases &s it decays to polonium-218, through the emission of alpha particles. The half life of radon-222 is 3-8 da ys, so after 38 days half the original amount is left; 7-6 days (two half lives! later, a quarter of the radon remains; after 11-4 days /three half lives), an eighth of the sample is left, and so on. The halflives of radioactive isotopes vary from fractions of a second to billions of years.

*• The Geiger counter measures levels of radioactivity - here it is being used to check the activity of produce at a market after the explosion of the nuclear reactor at Chernobyl. Invented in 1928 by the German physicist Hans Geiger and his colleague Walther Miiller, the counter consists of a gas-filled tube with a wire along its axis. When ionizing radiation passes through the gas, it liberates electrons which, under the influence of a high electric field in the tube, set off avalanches of electrons that are picked up by the central wire, and produce a signal.

radioactive gases they emit, for these can be inhaled, and inhalation like ingestion makes radiation more effective in ionizing critical parts of the human body. The main gas is radon, and its major isotope is radon-222, which again occurs in the decay chain of uranium-238. People living in areas where granite, for example, is common, are exposed to higher amounts of radiation from radon-222 in their houses than people living in other areas.

Natural radiation comes both from cosmic rays and from natural radioactivity. The best-known naturally-occurring radioactive element is uranium. Some 99 percent of uranium is in the form of uranium-238, the isotope (| page 66) that combines 146 neutrons with the standard 92 protons. It emits alpha particles, and it was in an experiment involving uranium salts that radioactivity was first discovered (4 page 79).

Radioactivity transmutes the atomic nuclei of an isotope of one element to those of an isotope of another element. Once all the nuclei of the original isotope have "decayed" in this way, that isotope no longer exists. The time this process takes varies from one isotope to another, because it depends on subtle balances of the changes in mass and energy involved in the transition. The process is also fundamentally random. The nuclei of a radioactive isotope do not all decay simultaneously, nor is it possible to say which nucleus is going to decay next. One nucleus may decay quickly, another may take a long time to decay. The rale of decay of a particular isotope is therefore characterized by a statistical quantity called the halflife. This is the time it takes for half the nuclei in a given sample to decay. The half-lives of the elements vary from fractions of a second for some isotopes, to millions of years for others and this provides a clue as to which isotopes occur naturally. At the time the Earth was formed some 4.5 billion years ago, it could have contained many radioactive isotopes. Those with halflives that are short compared with the age of the Earth will have long since decayed away completely. Only a few remain to this day, including uranium-238, with its halflife of 4-51 billion years. However, the continuing decays of uranium-238 give rise to nuclei that are themselves radioactive on shorter timescales. Indeed, uranium-238 is the start of a chain of decays by both alpha and beta emission. The chain passes through a dozen or so isotopes of different elements and stops eventually at a stable form of lead - lead-206. This chain includes radium-226, a far more powerful emitter of alphas than uranium.

The harmful effects of radiation are put to good use in the treatment of cancers

The units of radiation and radioactivity

The basic measure of radioactivity is the rate at which a substance decays. Historically the unit of radioactivity was defined in relation to the activity of one gram of radium, the radioactive elenei it that Pierre and Marie Curie discovered. ~or many years the basic unit was the curie (Ci), defined as 3.7x 10 ^{w}decays per second, approximately the rate from Ig of radium. Now the "unit of activity" is the becquerel (Bq), named for the discoverer of radioactivity, and defined as one decay per second.

The number of decays says something about the amount that there is of a particular radioactive substance, such as Ig of radium, or the relative activity of different substances, but it says nothing about the ionizing effects of the radiations produced. For gamma rays, and also for X-rays (which are not produced by radioactivity but which are akin to gamma rays), a unit is defined to quantify the ionizing effect of the radiation in air. This is called the "unit of exposure" and it is a measure of the amount ofX- or gamma radiation that produces enough ionization to create a total electric charge of one sign (negative or positive) of 1 coulomb in 1kg of air. Units called rontgens are often used to express exposure. One rontgen is equal to 2.58x 10$^{'4}$ coulombs per kilogram. These units reveal in a sense the amount of X-rays or gamma rays to which something or someone is exposed. A different unit is used to quantify the amount of radiation actually absorbed by a material. The "unit of absorbed dose" is called a gray (Gy), and it is defined as a deposited energy of 1 joule per kilogram. This is the energy the radiation gives up in ionizing the material. (The related unit, the rad, is equal to 0.01 Gy.)

For the purposes of calculating the amount of radiation that is safe for a human to receive, none of these units is satisfactory. This is because the different kinds of radiation have different effects on human tissues, and the importance of the effect depends critically on the part of the body that receives it. Radiobiologists work in "units of dose equivalent", called sieverts (Sv). These units are grays multiplied by a quality factor which takes into account the type of radiation and the long-term risks it has. Gamma rays and electrons have a quality factor of 1; for protons it is in the range 1 to 2, depending on the energy of the particles; for neutrons it ranges up to 10 or so for the most energetic particles; and for alpha particles and heavier ions, assuming that they are inhaled or ingested (because otherwise skin and clothing prevent any damage), the quality factor is as high as 20. Typical annual doses from natural radioactivity total in the region of 1 millisievert (one thousandth of a sievert). A lethal dose, which would kill within a month, is over 100,000 times greater.

A • High levels of radiation may not kill straight away, but "early effects" (occurring within a few weeks) include vomiting, secondary infections, and blistering or reddening of tha ekin larythema). Many victims of the nuclear hnmh dropped on Hiroshima during World War II showed such effects. "Late effects " refer mainly to an increased incidence of cancers appearing up to 30 years or more after exposure to radiation. Certain organs are more susceptible than others to the radioactive isotopes present in nuclear fallout.

T Different types of radiation have different potential for damaging tissue. Alpha particles are heavily ionizing and a few particles can deposit a significant amount of energy, as measured by the absorbed dose in grays (J/kg). A gamma ray, on the other hand, does not directly ionize atoms, but releases a few electrons which are only weakly ionizing.

Harmful isotopes

Iodine'3'

The body easily absorbs iodine either through the digestive system or through the lungs. The iodine quickly travels to the thyroid gland and can stay there several months. Radioactive iodine$^{13'}$ can be blocked from the thyroid by taking stable iodine pills.

Cesium$^{'34,,37}$

Muscles are the tissues (hat accumulate cesium and retain it for many months, Radioactive cesium^{137} is particularly problematic, with a halflife of 30 years. Both cesium^{137} and cesium^{134} enter the food chain via vegetation grown On contaminated soil.

Strontium^{90}

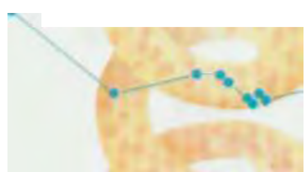
Bone cancer is one of the likely results of the absorption of too much Strontium^{90}. The strontium at first replaces calcium atoms nearthe surface of bones, but eventually it also leads to damage to the bone marrow itself, causing leukemia.

Carbon^{14}

Carbon^{14}, used for radiocarbon dating, is easily taken into the food chain in the form of carbon dioxide gas. Its halflife is very long-5,730 years - but fortunately it is quickly incorporated into carbon dioxide in the body, and is exhaled.

Effects of ionizing radiation

ofgtir 4* -^fikpyfi*~fi &



Alpha particle
1 gray = 4 tracks through cell nucleus

Gamma Ray
1 gray = 1000 tracks through cell nucleus

•elonised atoms

Chromatin fiber

Calculating radioactive dosage

Absorbed dose

A measure of the energy that radiation deposits in a material in the process of ionizing the atoms.

Dose equivalent

Absorbed dose multiplied by a quality factor to show variations in damage by different kinds of radiation.

Risk weighting factors

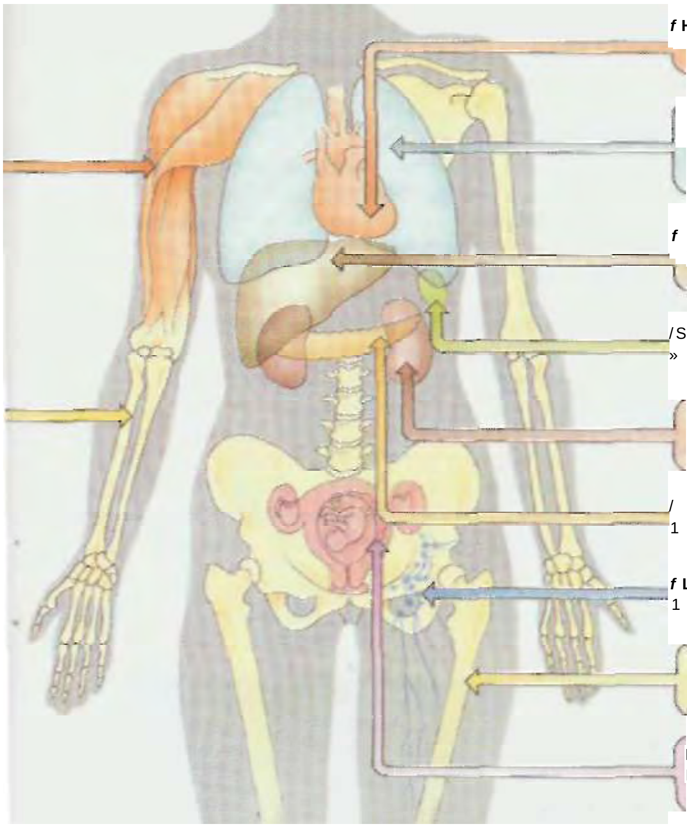
Testes and ovaries	0.25
Breast	0.15
Red bone marrow	0.12
Lung	0.03
Thyroid	0.03
Bone surfaces	0.03
Remainder	0.30
Whole body total	9.00

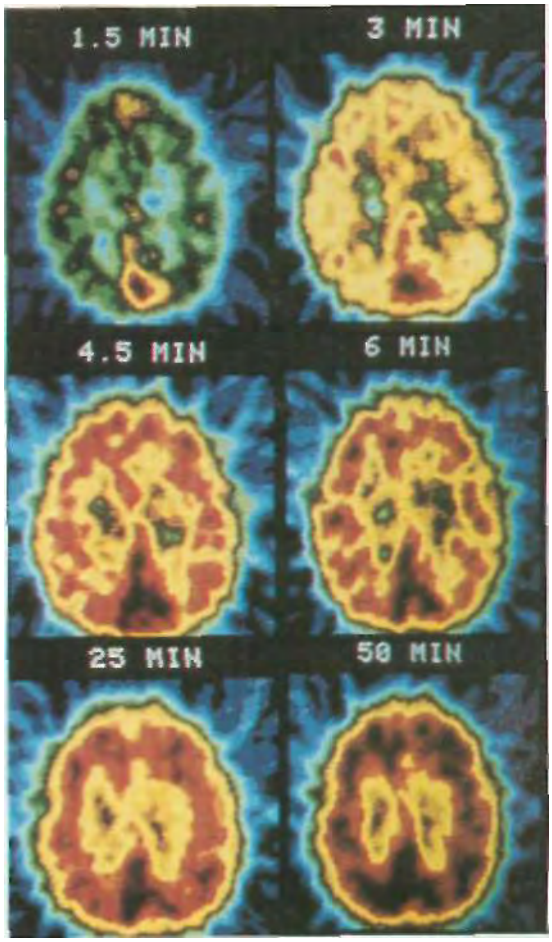
Effective dose equivalent

Dose equivalent multiplied by a risk weighting factor to account for the different susceplibilities of organs.

Collective effective dose equivalent

Average dose equivalent multiplied by the number of people in a population.





A Many kinds of radioactive isotope are made artificially for use in medicine. The various substances home in on different tissues and their radiations can be used either to produce images of organs or to damage tumor cells, in radiotherapy. The scans show views of the human brain revealed by the absorption of radioactively-traced glucose.

Biological effects

Ionizing radiations can be harmful to biological tissues; the more radiation a tissue is exposed to, the greater the likelihood of damage and the greater its severity. The damage to cells is done when the DNA (the molecule that contains the instructions for the cell's operation and reproduction) is broken. This occurs either directly, when the radiation itself breaks the molecular chain, or indirectly, by the attachment of free radicals /highly reactive groups of atoms) that the radiation has released elsewhere in the cell. Direct damage is usually caused by heavy charged particles, such as protons, alphas and heavier ions. Neutrons also give rise to direct damage, because they collide with nuclei in a cell, and produce nuclear fragments, including protons and alpha particles, which then break the DNA. Indirect damage generally occurs with electrons and high-energy protons, which ionize atoms in the cell and thereby release free radicals. X-rays and gamma rays also damage the DNA indirectly, because they energize atomic electrons within the cells, which then ionize other atoms. These harmful effects are put to good use in treating cancers.

Natural radiation accounts for far fewer deaths than, for example, air crashes or accidents in the home. Artificial radiations, from nuclear power plants, say, are equally low risk, provided adequate care is taken to keep the levels of radiation under control. Exposure to too much radiation causes cancers by damaging cells to the extent that they can still reproduce but no longer function effectively; still larger amounts of radiation produce severe effects, including burning, and can result in a swift death. Certain radioactive isotopes, such as strontium-90, iodine-131 and cesium-137, are particularly dangerous if they collect in specific parts of the body such as the bone marrow.

Artificial isotopes can also be extremely useful, especially in medicine where they are used as radioactive "labels" that reveal the way a particular substance is taken up in the body. Minute amounts of iodine-131 - which is dangerous only in large amounts - once injected into the body, will travel to the thyroid gland. By measuring the (temporary) radioactivity from this gland, doctors can discover if it is overactive or underactive, both of which are unpleasant conditions. Radioactive "tracers", such as the positron-emitters carbon-11 and oxygen-15, are used to image blood flow and oxygen metabolism in the brain. Positrons annihilate with electrons, producing gamma rays which are detected by a "gamma-camera" surrounding the head.

See also
Atoms and Elements 65-72
Studying the Nucleus 79-86
Nuclear Fission and Fusion 119-24

A Carol, aged eight, lives on Rongelap, an atoll of the Marshall Islands in the Pacific. She was deformed at birth, possibly as a result of US nuclear tests on the neighboring Bikini atoll in the 1950s. Data on longterm effects of radiation come only from such unfortunate cases.

T US servicemen cover their ears as they watch an atomic bomb explode in the Nevada desert in 1957, displaying the ignorance at that time about radiation.

The risks of low level radiation

One issue that worries many people is the risk from exposure to radiation in excess of the natural background level, either at work or in occupations that involve radiation, or accidently as in the case of the explosion at the nuclear reactor at Chernobyl in the Soviet Union in 1986. Particularly disconcerting for the nonspecialist are the wide variations in the risks that experts in different countries or different organizations claim to be safe.

After the accident at Chernobyl, the Soviet authorities suggested that an additional 42,500 deaths due to cancer would result from the fallout from the explosion; twelve months later they claimed that this figure was too high by10 or 20 times. In May 1987, experts in the United Kingdom concluded that there would be about 1,000 extra deaths in the European Economic Community (EEC) attributable to the accident, while a study in the United States put the figure ten or twenty times higher. All the figures are small in comparison with the total number of deaths from cancer, which will be about 30 million in the EEC alone in the 50 years following the accident. Yet the discrepancies can still be worrying.

Calculating the longterm risks from a radiation release such as this is a complex matter. It involves, for example, a knowledge of the composition of the fallout produced, an understanding of how the global atmospheric circulation might distribute the fallout, an understanding of how different kinds of radiation produce cancer, and so on. Each parameter in this calculation has its uncertainties, but one major problem and a major reason for scientific debate lies in estimating the risks due to the relatively low levels of radiation received by people away from the main region of the accident.

Very high levels kill, while somewhat lower levels cause cancers of various kinds. But it is hard to extrapolate from the known effects of high levels of radiation to calculate the number of deaths that might result from much lower levels. Much of the data scientists have at their disposal comes from specific groups of people - the survivors of the bombs dropped on Japan in 1945, people who lived in the Marshall Islands in the Pacific during bomb tests in the late 1940s and early 1950s, workers in uranium mines, workers using luminous paints (which contain radium) who licked their brushes to obtain fine points, people deliberately exposed to radiation in the course of therapy for a variety of ailments.

The basic and most common technique used to deal with this information is to develop a model for calculating the risk from radiation received by these specific groups, which in some cases applies to particular organs or tissues. The models should predict reasonably correctly the number of cancers observed in each group. The information from these models is then combined together to give an overall risk factor to the whole body. The arguments arise in extrapolating the results to • lower levels of radiation that the subjects in the various groups actually received. Some scientists question whether a simple linear extrapolation, where the risk remains directly proportional to the dose even at low levels, is valid. Monitoring people around Chernobyl will help resolve this question.

Nuclear Fission and Fusion

16

The energy stored in the nucleus...Splitting the uranium atom...Nuclear fission in peace and war...Energy from fusion... The force that fuels the Sun...pEBSPEcnvE... The race for fission... The Manhattan Project...A natural nuclear reactor...The missing solar neutrinos

Atoms are tiny storehouses of energy. Chemical reactions can release some of this energy by rearranging the way in which atoms combine in different substances. This happens, for example, when coal burns and in the spectacular explosions of gunpowder and fireworks. Reactions such as these involve changes in the energy of the electrons within atoms and molecules.

However far more energy is released when an atomic nucleus rearranges itself. The energies binding the nucleons (protons and neutrons) together in a nucleus (| page 82) are typically a million times greater than those that bind the electrons in atoms and molecules (4 page 57). Thus the energy required for and released by nuclear reactions is a million times greater than the energies involved in chemical reactions. In other words, roughly the same energy can be produced from nuclear reactions in one gram of uranium as from the burning of a million grams (1 tonne) of coal.

The best known methods for a nucleus to release energy are fission (splitting into two smaller fragments) and fusion (joining two nuclei together). Another way that a nucleus can release energy is through radioactivity (| page 82). This occurs in nuclei that can change to a configuration of nucleons with a lower total energy, for example by emitting a cluster of two protons and two neutrons (alpha decay), or by converting a neutron into a proton (beta decay). For any total number of nucleons - represented by the so-called mass number, A - the rearrangement with the least energy is the most stable, and gives the most common form of nucleus for a particular element.

The discovery of fission

Nuclear fission was discovered in 1938 by two German physicists. Otto Hahn (1879-1962) and Fritz Strassmann (b. 1902). They were bombarding uranium with neutrons when they identified barium nuclei among the fragments from the collisions. Barium, with 56 protons, is much smaller than uranium, which has 92 protons.

Two of the researchers' colleagues, who had fled from the Nazis in Germany, Use Meitner (1878-1968) and her nephew Otto Frisch (1904-1979) who was working in Copenhagen in Denmark, suggested that a neutron could cleave a uranium nucleus in two - a process they named fission. Meitner and Frisch tested their idea experimentally and proved that the uranium nucleus does break into two more or less equal parts, such as barium and krypton (with 47 protons).

The Danish theorist Niels Bohr (1885-1962) leapt upon this idea, and together with the American John Wheeler (b. 1911), he developed the liquid drop model of the nucleus, which describes several properties of atomic nuclei (fpage 86). One feature of neutron-induced fission soon became apparent. It released neutrons which could produce more fissions and so establish a chain reaction.

In September 1939 Germany invaded Poland and World War II began. The possibility of harnessing the energy released in a fission chain reaction to make a powerful bomb was not lost on scientists outside Germany, especially on those who had fled the Nazis. In the United States, physicists urged the government to support research in nuclear physics as a matter of urgency. But it was not until the end of 1941 that a concerted effort on establishing the possibility of a chain reaction really began.

The task fell to Italian physicist Enrico Fermi (1901-1954). In December 1941 he was summoned to Chicago, where American physicist Arthur Compton (1892-1962) was head of the "uranium project". A year later Fermi's team had built the first "atomic pile" and on 2 December 1942 it achieved the first artificial chain reaction.

• The Italian physicist Enrico Fermi, who had emigrated to the United States and was awarded the Nobel Prize for physics in 1938 for his work on radioactivity, built the world's first atomic pile underneath the squash courts at the University of Chicago. On 2 December 1942 his team produced the first sustained chain reaction on it.

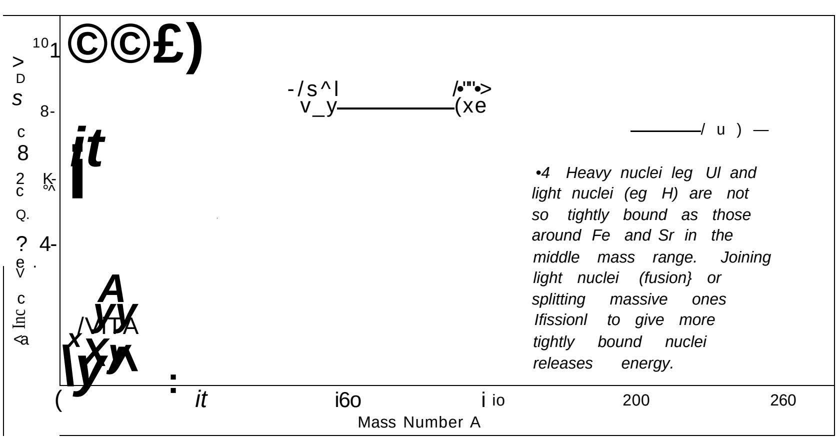
The world's first nuclear reactor occurred naturally almost 2,000 million years ago

One way to compare the stability of the nuclei of different elements is to calculate the difference between the mass of each nucleus and the total mass of the individual nucleons it contains (| page 92). In general, the commonest (most stable) nuclei of each element have a mass that is less than the total mass of the constituents. The effect is greatest for medium-sized nuclei, from about strontium (A=90) to cerium (A=140). It is smallest for the lightest nuclei (A less than 20) and the heaviest nuclei (A greater than 210).

The total mass of two medium-sized nuclei is *less* than the mass of a nucleus of double the size. Thus when a large nucleus is split in two, a net reduction in the total mass occurs, and the amount by which the mass changes appears as energy (• page 43). This process is known as nuclear fission, and the way in which it occurs can be understood by regarding the original heavy nucleus as a drop of liquid.

If a drop of liquid becomes sufficiently distorted from its natural spherical shape, it splits into two drops, because the two smaller spherical drops require less energy than the distorted large drop. A similar process occurs if a large nucleus becomes distorted. This can happen spontaneously in some nuclei because an effect of quantum wave mechanics (| page 102) allows the nucleus to distort. But such "spontaneous fission" is very rare. For example, in the common form of uranium, uranium-238, the halflife (| page 115) due to spontaneous fission is 8 x 10^{15} years. However, nuclear fission can be made to occur by bombarding appropriate nuclei with neutrons. A nucleus of uranium-238 will capture an energetic neutron and form an excited (energized) nucleus of uranium-239. This new nucleus distorts far more easily, and within a fraction of a second - about 10~~ seconds - it splits into two smaller fragments.

The two nuclei produced in fission are usually accompanied by a few much smaller pieces, including neutrons. A large nucleus contains a greater proportion of neutrons than a smaller nucleus, in order to dilute the repulsion between the many electrically-charged protons. Thus the two smaller nuclei produced in fission contain too many neutrons, and these are shed almost immediately after the fission occurs. In a sample of fissionable material these neutrons can induce further fissions, thereby producing a self-sustaining chain reaction with a release of energy at each step in the chain. If controlled, such a chain reaction yields a steady flow of energy, the principle behind the operation of nuclear reactors. If the chain reaction is uncontrolled, it can lead quickly to a catastrophically huge release of energy, and the result is an "atom" bomb.



•4 Heavy nuclei leg UI and light nuclei (eg H) are not so tightly bound as those around Fe and Sr in the middle mass range. Joining light nuclei (fusion} or splitting massive ones Ifissionl to give more tightly bound nuclei releases energy.

A M Robert Oppenheimer, (left) and General Leslie Groves survey the remains of the tower on which the world's first nuclear explosion had occurred on 16 July 1945- the "Trinity" test. The explosion in the desert at Jornada del Muerto, near Alamogordo in New Mexico, had crystallized the earth at the base of the tower. Groves, who was in charge of the project to build the bomb, had recognized the need for a special laboratory and had appointed Oppenheimer, one of the world's most brilliant theoretical physicists, as its director.

The father of the atomic bomb

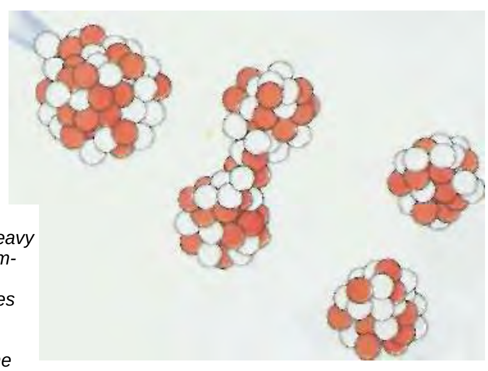
In November, 1942, General Leslie R. Groves of the US Army visited Los Alamos, a remote area of New Mexico. By the end of World War II, the site had become home to some 6,000 people, all working on the Manhattan Project - America's effort to build an atom bomb. The director of the Los Alamos laboratory was J. Robert Oppenheimer (1904-1967), a theoretical physicist who has become one of the most fascinating figures of 20th-century science.

Oppenheimer headed the laboratory at Los Alamos with great flair, collecting about him a team of scientists whose brilliance and dedication remains unparalleled. Their work came to fruition on 16 July 1945, with the "Trinity Test", the explosion of the world's first atomic bomb, in the desert in New Mexico. At this stage, Oppenheimer and his fellow scientists were jubilant with success; later, after the horrors of the bombs dropped on Hiroshima and Nagasaki became apparent, many of the same scientists became concerned about the spread of these weapons. Oppenheimer himself opposed the development of the hydrogen bomb, and in 1954 he was declared a security risk and banned from government work. Only in 1963 did the government make amends by awarding Oppenheimer the Fermi medal for his work.

A natural nuclear reactor

In the 1970s, scientists measuring samples of uranium ore from a quarry in Gabon, West Africa, found evidence of a chain reaction that occurred 1,780 million years previously. The Oklo quarry is evidently the site of a natural nuclear reactor that was active in the Precambrian geological era.

Today, uranium-235 normally forms 0-72 percent of naturally-occurring uranium, a percentage that is fixed by the relative half lives of the two isotopes. The Okla sample, however, contained unusually low proportions of uranium-235, especially where the ore was rich in uranium. Traces were found of other rare elements in proportions typical of artificial nuclear reactors. Presumably, when the proportion of uranium-235 was as high as 3 percent, there was enough fissile uranium in the rock for the fission process to go critical and a chain reaction to begin. Water trapped in the surrounding sandstone acted as a moderator, and the chain reaction continued for almost a million years.



A In nuclear fission, a heavy nucleus, such as uranium-238, absorbs an extra neutron. This destabilizes the nucleus so that it distorts and splits into two smaller nuclei, at the same time releasing a few neutrons. These neutrons can be used to induce more fissions in a chain reaction.

c

Nuclear fusion is the process which fuels the Sun and other stars

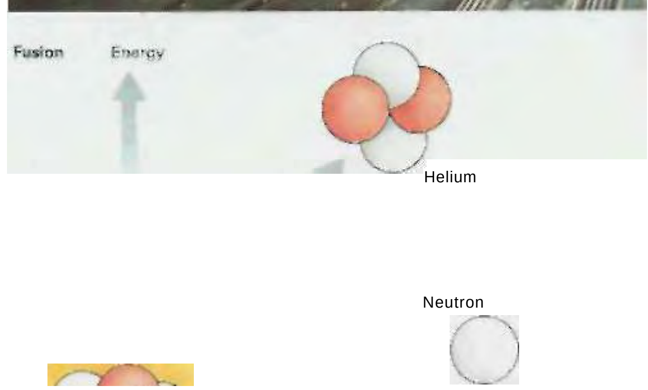
Nuclear fusion

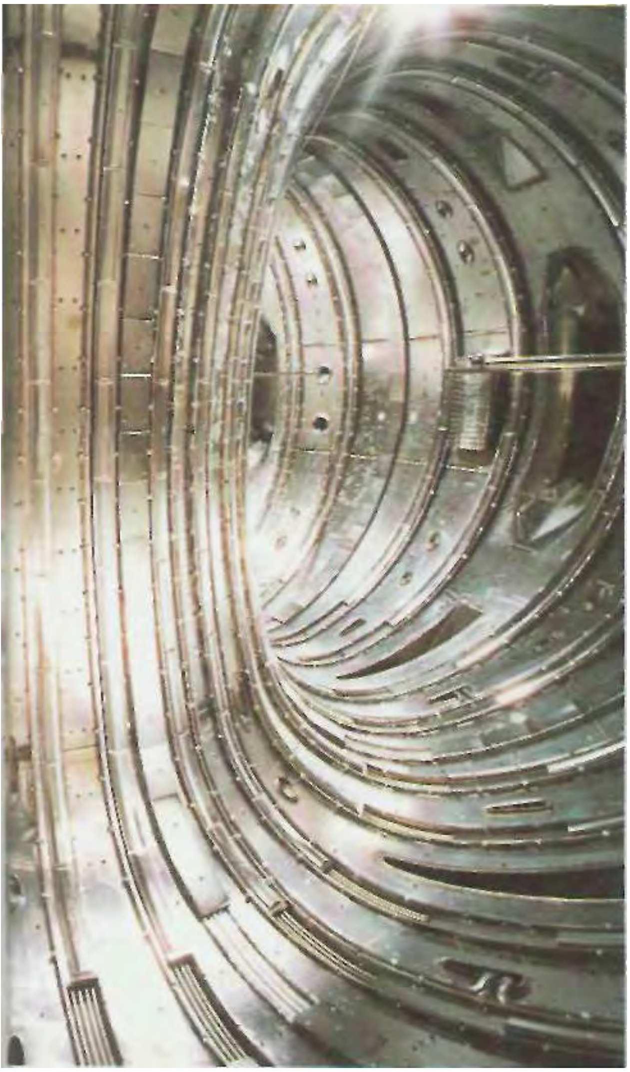
An alternative way to release the energy contained in atomic nuclei involves light nuclei such as hydrogen and helium. In this case the total mass of two light nuclei is *greater* than the mass of a nucleus of double the size. Thus combining two light nuclei to form a heavier nucleus gives a net reduction in mass, and is accompanied by a release in energy to keep the total balance of mass-energy constant. This type of reaction is known as nuclear fusion, and it is the process that fuels the Sun and other stars.

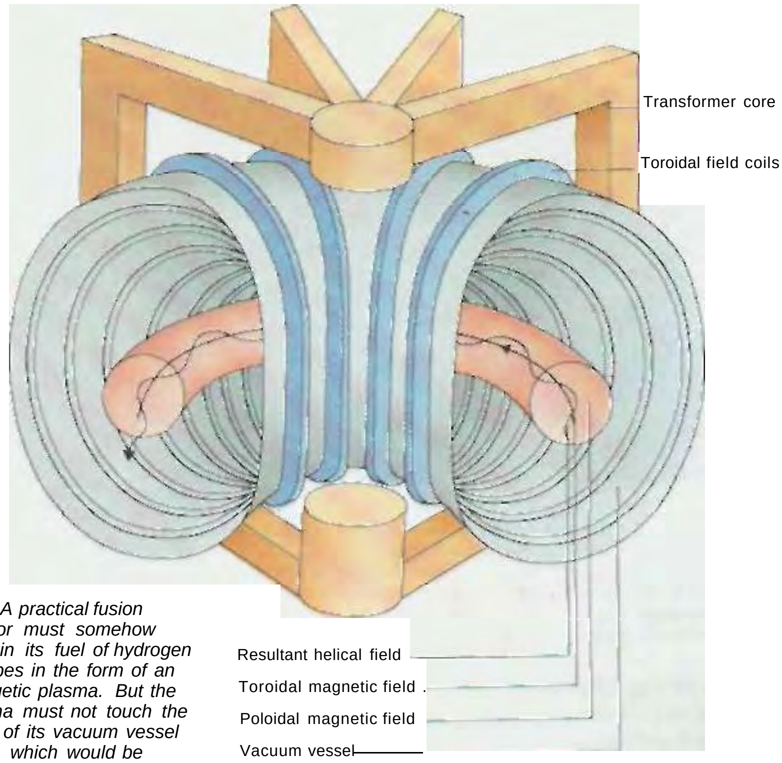
Physicists have been trying for several decades to produce energy from hydrogen fusion here on Earth. They have been successful in making "hydrogen" bombs, which release fusion energy in a sudden, violent manner, but they have still to achieve the production of energy in controlled conditions. The main problem in inducing nuclear fusion lies in overcoming the natural electrical repulsion of the like-charged protons. This happens in the center of the Sun because matter there is a hundred times more dense than water, and at temperatures of 15 to 20 million degrees the particles in the core collide very energetically. Even so, fusion reactions in the Sun are relatively infrequent. A proton will exist in the core on average for 10 billion years before it joins with another to release deuterium; such is the weakness of the weak nuclear force responsible for the basic reaction. Physicists need to create high temperatures, like those in the Sun, and very high densities if they are to achieve practical rates in fusion reactions here on Earth. They also intend to "improve" upon the Sun by using reactions that occur millions of times faster than the Sun's basic proton-proton reaction (^ page 124). The fusion of deuterium nuclei, either with each other or with tritium nuclei, offers the best opportunities. These materials are heavy isotopes of hydrogen - that is, they are forms with the same number of protons (one) but with an additional one or two neutrons in deuterium and tritium respectively.

To achieve suitable temperatures and densities in a fusion reactor the "fuel" of deuterium and tritium must be in the form of a plasma. This is a state of matter in which electrons and nuclei are completely separated: it is a totally ionized gas. A major problem lies in containing such a plasma, which must not be allowed to touch conventional walls. The favored option is to use magnetic fields because they can exert a force on the charged particles of the plasma.

Fusion offers the possibility of an inexpensive alternative to fission for releasing nuclear power, because naturally-occurring deuterium could be extracted from sea-water. It also promises to be inherently safer than fission power, which is based on a self-sustaining chain reaction that can in principle surge out of control. However, as physicists struggle to understand the behavior of plasmas, the practical application of fusion power still seems to lie far in the future.







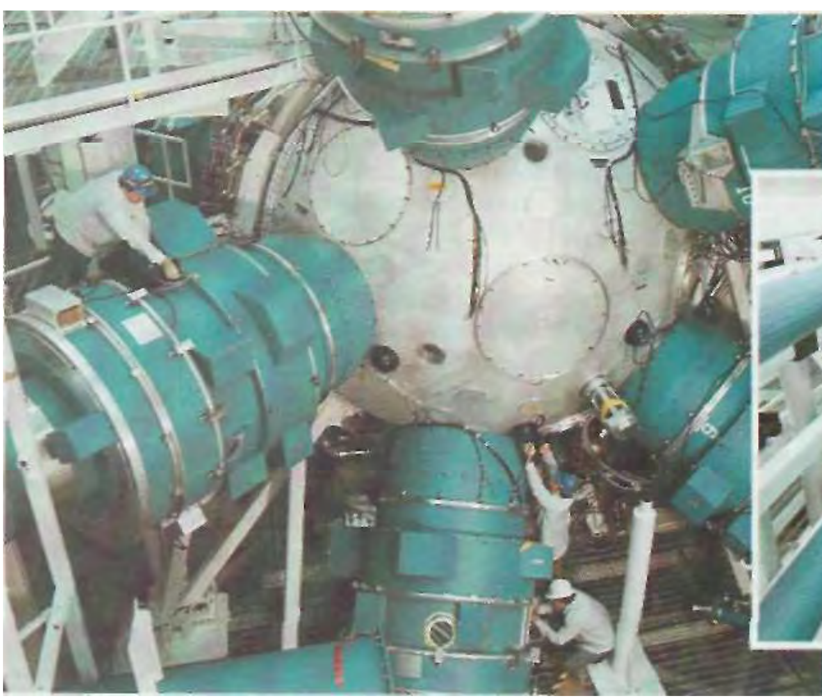
•4 A A practical fusion reactor must somehow contain its fuel of hydrogen isotopes in the form of an energetic plasma. But the plasma must not touch the walls of its vacuum vessel (left), which would be vaporized. One solution - the tokamak- is to contain the plasma within a magnetic field shaped like a torus. The magnetic field is set up by D-shaped coils around the vacuum vessel.

• • An experiment in laser fusion at the Lawrence Livermore laboratory in California has a complex of massive lasers to deliver thousands of joules of energy for tiny fractions of a second to a deuterium-tritium pellet target, smaller than a grain of sand.

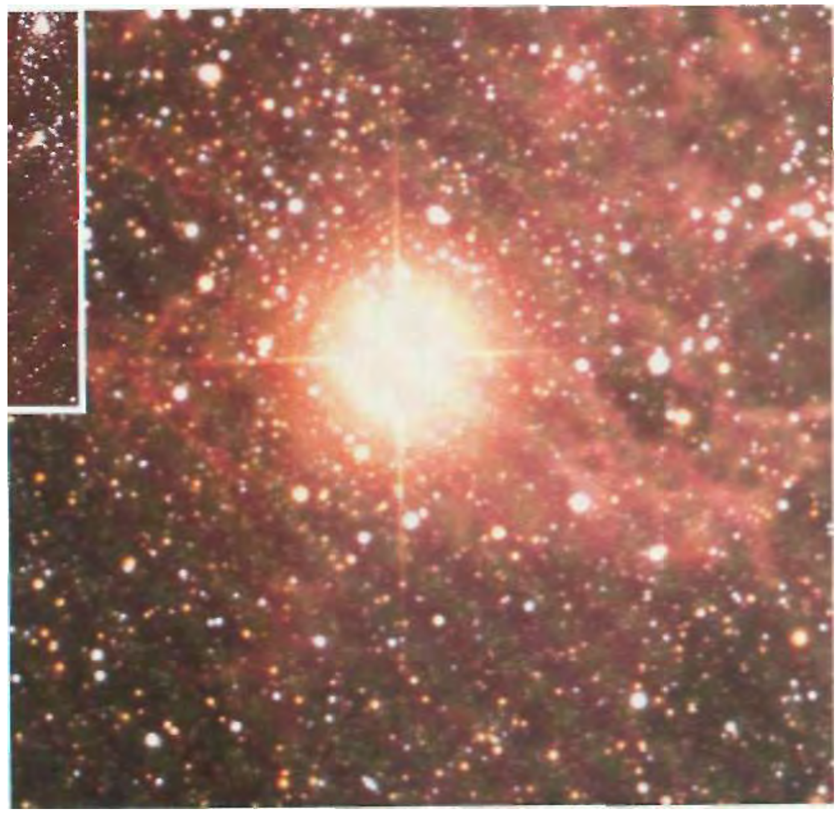
Laser fusion

*One alternative to the magnetic confinement of fusion fuel is the approach known as inertial confinement. This relies on the inertia of the fuel to keep it together once it has been compressed to **a** very high density. The aim is to use many beams of particles or laser light to bombard a glass-walled pellet of fuel so as to generate the conditions to "bum " ail the fuel. The beams would evaporate away the outer layers of the pellet, inducing the remainder to implode. The implosion would compress the fuel and could lead to fusion.*

Most research on inertial confinement fusion has so far been done with powerful laser beams directed at pellets a few tenths of a millimeter across. For a useful power station, pellets would have to be several millimeters across and the lasers used would have to deliver many megajoules of energy over periods of several hundredths of a microsecond.



See also
Studying ihe Nucleus 79-86
Radiation and Radioactivity 111-18



A •• *Heavy elements, particularly those heavier than lead, are probably produced in processes that occur in a supernova - a stellar explosion. Supernova SN1978a (right), observed in 1987, was the first nearby supernova to be seen for nearly 400 years. The image above shows the same region of sky before the explosion.*

The missing solar neutrinos

Several steps in the chain of reactions that fuel the Sun emit particles called neutrinos. These particles have little or no mass (certainly less than one ten-thousandth the mass of the electron), no electric charge, and they interact only via the weak nuclear force. This means that, although it has been estimated that every cubic centimeter of space contains between 100 and 1,000 neutrinos, they interact only very rarely with matter. Indeed, neutrinos from the Sun can pass right though the Earth.

However, it is still possible to detect the very rare interactions. All that is needed Is a large detector built from a suitable material, and a reasonable amount of patience. One such detector exists 1,500m below ground in the Homestake gold mine in South Dakota. Here United States physicist Ray Davis, of the Brookhaven National Laboratory, has been picking up neutrinos from the Sun since the mid 1960s. His detector consists of 400,000 liters of perchlorethylene (dry-cleaning fluid). The neutrinos interact with chlorine in the liquid to produce a radioactive form of argon-37. Davis flushes the argon out of the tank periodically and measures the amount of radioactivity. He can then work out how many solar neutrinos his detector has intercepted.

The surprise is that the experiment consistently detects only one third the number of neutrinos that are expected according to theory. No one can find fault with the experiment, nor with the theory of the Sun's interior. Possibly the neutrinos have a small mass. Another intriguing possibility concerns the nature of the neutrinos themselves. Do some of them change character from being so-called electron-neutrinos, which the Sun emits, to become one of the other types that particle physicists know exist (4page 102)7 Davis has not been able to detect other kinds of neutrino. The answer to this question is one of the challenges facing particle physicists, and has given rise to a number of new kinds of solar neutrino detector with which physicists hope to cast new light on the problem during the 1990s.

Fusion in the heart of the Sun

The basic process that fires the Sun's furnace is the fusion of four hydrogen nuclei (protons) into a nucleus of helium-4 (which consists of two protons and two neutrons). For this to occur, two of the four protons must somehow convert into neutrons. This happens each time two protons combine to form a nucleus of deuterium, which contains a proton and a neutron. This reaction is in a sense the opposite to the beta decay of a neutron to a proton. Like beta decay it occurs via the weak nuclear force, and is accompanied by the emission of a positron (positively-charged antieleclron) and a neutral particle known as the neutrino. The deuterium produced in this way combines with another proton to form a nucleus of helium-3, and this in turn reacts with another nucleus of helium-3 to form helium-4 and two protons. The net result of this "proton-proton chain" - named after the initiating reaction - is the formation of helium-4 from hydrogen, accompanied by a release of energy which is all-important for life on Earth.

Once there is sufficient helium in a star it can form heavier nuclei through additional fusion reactions, leading quickly to the production of carbon-12. Carbon-12 contains six protons and six neutrons and is therefore directly equivalent to three nuclei of helium-4. The simplest way that it forms in stars is for two helium-4 to fuse, making beryllium-8 which then fuses with a third helium-4 nucleus to make carbon-12. The carbon-12 reacts with hydrogen initiating a sequence of events that has the same net result as the proton-proton chain - namely, the formation of helium-4 from four protons, with the attendant release of energy. During this chain of reactions the carbon-12 is converted into nitrogen and then to oxygen, which then reverts back to nitrogen and finally to carbon-12. Thus the carbon is not used, but acts as a catalyst for the hydrogen "burning". This carbon-nitrogen cycle is important in stars that are hotter than the Sun.

Index

Page numbers in *italics* refer to illustrations, captions or side text.

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Credits

Key to abbreviations: ARPL Ann Ronan Picture Library; CD Chemical Designs Ltd.; RHPL Robert Harding Picture Library; SPL Science Photo Library; b bottom; bl bottom left; br bottom right; c center; cl center left; cr center right; t top; tl top left; tr top right: l left; r right

Cover SPL/Gordon Garradd I SPL/Martin Dohrn 2-3 RHPL. A.Carr 4 CF.RN 5 SPL/Fred Espanak 6 John Hillelson Agency.'R. Michaud 7,8tl BPCC'Aldus Archive 8tc RHPI. 8-9 CD 11 Tony Stone Worldwide 12ir Leo Mason 12-13 RIIPL 13tl Colorsport 13tr Rex Features 14-15 Allsport 14b National Maritime Museum 15bl Biofotos/H. Angel 16 Mountain Camera/John Clearc 17 SPL. Jan Hinsch 18tl ZEI-'A'L. Vilotta 18cr Richard House 18b Bridgcman Art Library/Royal Society 19

© David Redfern/Stcphen Morley 19r ZEFA/F. Damm 20 Michael Freeman 21t ARPL 21b SPL/Dr Gary Settles 22t John S. Shelton 22b SPI./Hank Morgan 23 ZEFA/Justitz 24 MOD (PE) RAE Farnborough/Crown Copyright 25 ARPL 26bl SPL/Claude Nuridsany & Marie Perennou 26cr David Lee Photography 26-7 Hutchison Library 27c SPL/ Manfred Kagc 27br Allsport 28l SPL/Andrew McClenaghan 28b Burndy Library 28-9 J. Allan Cash 29br© David Redfcrn'Andrcw Puller 30t Bruce Bennett 30b Mountain CamcraColin Monteath 31l Mountain Camera/John Cleare 31b Marin 32t Sonia Halliday & Laura Lushington Photographs 32-3 National Physical Laboratory/Crown Copyright 33cl SPL/James Bell 33r Bruce Coleman Ltd/Stephen J. Krascman 34t Paul Brierley 34h John Watncy Photo Library/Jerry Young 35 High Altitude Observatory and National Center for Atmospheric Research, Colorado 36cl Nils Abramson 36-7 John Watncy 37r Perkin Elmer 38cr Royal Society 39t Tony Stone Worldwide 40 SPL/Martin Dohrn 41 SPUJeremy Burgess 42 SPL/Fermi National Accelerator Lab 43 Krench Railways Ltd 44 Paul Brierley 45t SPL/Alex Bartcl45b SPL'Vaughan Fleming 46b Michael Holford 46-7 SPL/Jack Finch 48t RHPL 48b Dr Kenishead 49t SPL'Lawrence Berkeley Lab 49b ARPL 50-1 ZEFA'T. Ives 51t British Library 51b SPL/James Stevenson 52t Paul Brierley 52b SPL/DavidTaylor 54 SPL/David Scharf 55l Paul Brierley 55b Minolta 56 RHPL 57 BBC Hulton Picture Library 58t Science Museum, London 58b Tim Woodcock 59 Tony Stone Worldwide 60ll Patrick Moore 60c National Maritime Museum 60bl SPL/Dr W.H. Ku 60br SPL/NASA 61tl Institution of Electrical Engineers 61tr ARPL 61bl SPL/NASA 61br SPL/Max Planck Institute for Radio Astronomy 62-3 Nikola Tesla Museum Belgrade 63insei Nikola Teste Museum. M. Connolly Collection 63bl Independent Television News Lid 64 David Rcdfern 65 ARPL 66l Bridgcman Art Library/Musee des Beaux Arts, Lille 66br SPL/Dr MiisuoChisuki 67 London Pictures Service/Crown Copyright 68 University of Cambridge, Cavendish Laboratory 69 CI) 71 Kovosti Press Agency 731 Lawrence Clarke 73bl,br Dr F.B. Pickering 74lr Leo Mason 74br Hoechst High Chem 74bl Frank Spooner Pictures 75tl Andrew Lawson 75br Michael Freeman 75hl Institute of Archaeology, University of London 76 ZEFA/W.H. Mueller 77tl Susan Griggs Agency 77bl Vamier de Nanxc 77br W.K. Young, National Meteorological Office 78ll SPLllauk Morgan 78tr SPL.'Swindon Silicon Systems/A. Sternberg 78br Art Directors 791 SPL/NASA 79insel AIP Niels Bohr Library, William G. Myers Collection 79b Mansell Collection 80i University of Manchester, Physics Dept 80b University of Cambridge, Cavendish Laboratory 81 SPL'Lawrence Berkeley I .aboratory 82 SPL/David Paker 83 Topham Picture Library 86t Niels Bohr Institute, AIP Niels Bohr Library 86b,87 Bibliothck und Archiv zur Geschichtc der Max Planck Gesellschafl 88 SPL/Physics Dept, Imperial College 89 SPL/Hank Morgan/University of Massachusetts at Amherst 90-1 SPL/David Scharf 91 Bell Labs 92t CD 92-3 SPLCNRI 94t Pete Addis/New Scientist 94b J. F. Allen, St Andrews University Physics Depl 94-5 CD 95 J.F. Allen, St Andrews University Physics Depl 96t Allsport 96b Osterreichische National Bibliothek 98 SPL/Lawrence Berkeley Lab 99 main picture SPL/Stanford Linear Accelerator Center 99tr CERN 99br5PL/Tasso Experiment, Desy/Oxford Nuclear Physics Lab 100,101 SPL/Lawrence Berkeley Lab 102ir SPI ./Powell Fowler & Perkins, "The Study of Elementary Particles" 102-3 SPL-'Irvine-Michigan-Brookhaven Prolon Decay Experiment 104,1061 SPL/David Parker 106-7 SPL. Patrice Loiez, CERN 1071 SPL/Brookhaven National Lab 109 University of Manchester, Jodrell Bank 110 SPLIrviiic-Michigan-Brookhavcn Proton Decay Experiment 111 Popperfoto 112tl SPL'C.T.R. Wilson 112tc SPL 112tr SPL1. Jolioi Curie & F. Joliot 112bl,br Popperfoto 113t SPOOhio Nuclear Corporal ion H3bi SPL/Martin Dohrn 114bl Susan Griggs Agency 114-15 British Museum 115ir Frank Spooner Pictures 116 Camera Press 117 SPL/Cca Orsay/CNRI 118t Greenpeace' Pereira 118b TRH'IIS Army 119 Chicago Historical Society 120-1SPL/US Army 121t Topham Picture Library 122-3 JETJoinl Undertaking 123br,l23bl Lawrence Livermore Lab 124 Anglo-Australian Telescope Board

Artists
Robert and Rhoda Burns, Simon Driver, Chris Forsey, Alan Hbllingbery, Kevin Maddison, Colin Salmon, Mick Saunders

Indcxer
Barbara James

Typesetting
Peter MacDonald/Ron Barrow

Production
Joanna Turner