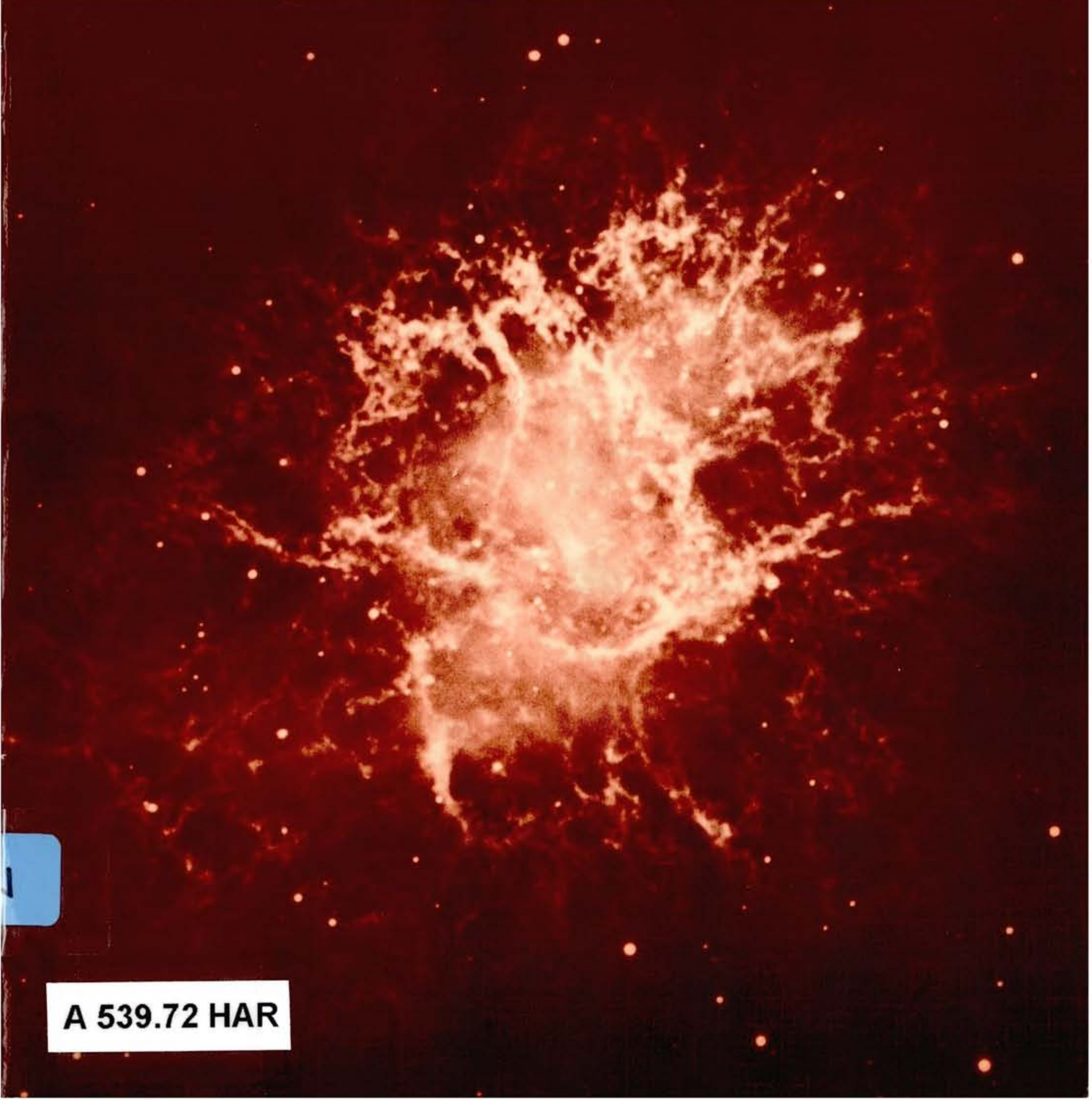
REVISED

NUFFIELD ADVANCED SCIENCE

# PHYSICS

## PARTICLES, IMAGING, AND NUCLEI



General editor,
Revised Nuffield Advanced Physics
**John Harris**

Consultant editor
**E. J. Wenham**

Editor of this book
**John Harris**

**Contributors**
Paul Davies
Derek Eastham
Hester Greenstock
George Marx
Charles Taylor



**Longman Group Limited**
Longman House, Burnt Mill, Harlow, Essex CM20 2JE, England
and Associated Companies throughout the World



Design and art direction by Ivan Dodd
Illustrations by Oxford Illustrators Limited

Filmset in Times Roman and Univers
Printed in Great Britain by Scotprint Limited, Musselburgh, Scotland

ISBN 0 582 35420 X



**Cover**
The Crab Nebula, the remains of a star (supernova) whose explosion in 1054 AD was recorded by the Chinese. It is a continually expanding mass of gas and high energy particles emitting radio waves and X-rays as well as visible light. Supernovae are discussed in the articles 'The particles and forces of nature' and 'Our nuclear history'.
*Mount Wilson and Palomar Observatory.*

Copy 1

# PHYSICS
# PARTICLES, IMAGING, AND NUCLEI

**REVISED NUFFIELD ADVANCED SCIENCE**
Published for the Nuffield–Chelsea Curriculum Trust
by Longman Group Limited

# CONTENTS

# The particles and forces of nature

**PAUL DAVIES**
**Department of Theoretical Physics, University of Newcastle-upon-Tyne**

## The key to the Universe

Over two millenia ago Mankind embarked upon a monumental quest. Its goal was to discover the ultimate building blocks of all matter, and to understand the natural forces that control and structure the material world. Many physicists believe that today we are within sight of this goal. Their optimism stems from the astonishing discoveries made in recent years about the world within the atom.

The sub-atomic domain is entered using gargantuan machines called particle accelerators. These enable fragments of matter to collide with enormous energies, at speeds very close to that of light. These accelerators can be many kilometres long, and usually consist of a ring-shaped vacuum tube in which charged particles such as electrons and protons are accelerated to ever greater energies. The paths of the circulating particles are bent by powerful magnets. On impact, new forces and particles are released, which can be studied by devices such as bubble chambers.

The microcosmos revealed by these experiments is inhabited by weird splinters of matter. Many of the particles created by high-energy impacts are violently unstable and exist only fleetingly, for minute fractions of a second, before decaying into more stable forms. A major challenge facing the physicist is to catalogue the properties of this rich variety of forms, and to understand how the many different particles relate to each other.

The detailed behaviour of all subatomic systems is governed by the laws of quantum physics, which endow the particles with some strange and unfamiliar properties. Fortunately, however, the broad features of the subject can be described without recourse to advanced quantum concepts. Another branch of the new physics directly relevant to these studies is the Theory of Relativity. Moving at close to the speed of light, the particles are subject to extreme relativistic effects, such as time dilation and variation in mass. Such effects must be taken into account in the design and interpretation of particle physics experiments.

Thus high-energy subatomic particle physics is the great testing ground of the new physics, and represents one of the most important frontiers of modern fundamental science. It commands the lion's share of scientific resources and Nobel prizes. New discoveries are being made all the time, and we can look forward to major new advances in the near future.

## The four forces of nature

In spite of the hugely varied and complex array of physical phenomena in the Universe, physicists recognize just four fundamental forces which control all natural activity.

### Gravity

The force of gravity is familiar in daily life, and was the first force to receive a systematic scientific description – by Newton. Gravity is a universal force; it acts between all particles of matter and between all forms of energy. It is a purely attractive force, and in a real sense it binds the Universe together. Although the force of gravity due to a point source diminishes with distance (according to Newton's inverse-square law, force $\propto 1/r^2$), its effects can be felt at great distances across space. For example, the Sun's gravity holds the Earth in its orbit.

The ability of gravity to act at a distance across empty space is explained using the field concept. Each source of gravity, for example a particle of matter, generates an invisible gravitational field which emanates from the source into the surrounding space. Another particle, immersed in this field, experiences a force. Viewed this way, the field can be regarded as a mechanism whereby one object can signal its presence to another at a distance, and command it to respond to its mass.

Gravity is by far the weakest of the four forces. Its effects are negligible among subatomic particles, and are hard to measure even for objects of everyday size. Only in astronomical bodies do the cumulative gravitational effects of enormous numbers of particles enable gravity to dominate the other forces.

### Electromagnetism

Electricity and magnetism were familiar to the Ancient Greeks, but it was not until the nineteenth century that their true relationship was established. The work of Faraday and others demonstrated that electric and magnetic forces are closely interwoven. For example, electric currents produce magnetic fields, while changing magnetic fields can induce electric currents to flow. Maxwell succeeded in formulating a unified *electromagnetic* field theory, in which charged particles and currents act as the sources of the field. One consequence of Maxwell's theory is that wavelike disturbances can propagate through the electromagnetic field. Light and radio waves are examples of such electromagnetic waves.

The electromagnetic force differs from gravity in a number of respects. Firstly, it is not universal; it only 'couples' to electrically charged particles, whereas all particles couple to gravity. Secondly, electromagnetism is a much stronger force. The ratio of electromagnetic to gravitational forces in an atom is $10^{39}$.

### The strong force

When the structure of the atom began to be understood this century it became clear that new types of force would have to be postulated to explain the processes that occur in the atomic nucleus. Evidently a strong attractive force must be present in the nucleus to bind the protons and neutrons together, especially as the former experience a powerful electrostatic repulsion due to their electric charge.

Experiments show that the strong nuclear force is very short in range, dwindling rapidly to zero beyond about $10^{-14}$ m. Its effects cannot therefore be experienced directly in the macroscopic world, but are confined to sub-atomic regions. This is in contrast to the long-range nature of gravity and electromagnetism.

The strong force is, as its name suggests, the strongest of all the forces of nature, but its true character has only recently become apparent. The actual forces within an atomic nucleus are immensely complicated, and for many years physicists had no proper theory with which to describe them. In the last decade or so, however, all this has changed with the realization that protons and neutrons are not *elementary* particles at all, but composite bodies. It is now accepted that these nuclear particles are themselves made up of smaller entities called quarks. The nuclear force which acts between protons and neutrons is seen to be only a vestige of the much more powerful interquark force that exists *within* the nuclear particles.

The strong force acts only between quarks. Other particles – those not made up of quarks – do not experience the strong force.

### The weak force

When the phenomenon of radioactivity was properly discerned in the 1930s, physicists were compelled to postulate a fourth type of force, called the weak force. It too is a nuclear force, but it is usually much weaker than either the strong force or electromagnetism.

The weak force manifests its presence chiefly by promoting the decay of unstable particles. The most familiar example of the weak force in action is the disintegration of the neutron. An isolated neutron will, after several minutes, abruptly disappear, to be replaced by a proton, an electron, and an unseen particle called a neutrino. This process is driven by the weak force. It can also act to cause particles to scatter from each other.

The range of the weak force is exceedingly limited – to about $10^{-17}$ m – so, like the strong force, it is confined to the deepest recesses of matter. Most particles are subject to the weak force, but in many cases its effects are swamped by those of the electromagnetic or strong forces.

Although the weak and strong forces can only be discerned directly on a subatomic scale, they are ultimately responsible for important large-scale phenomena, for example, sunlight and nuclear power. The weak force also plays a role in the spectacular and cataclysmic destruction of heavy stars in explosive events known as supernovae.

## Particles

Since the simple picture of matter as composed of electrons, protons, and neutrons was developed physicists have discovered – and postulated – many other particles. It is useful to group today's fundamental particles into two classes.

### Leptons

Of all the particles known to physicists, the members of one class possess especially simple properties. These are the leptons, meaning 'light ones'. At present five leptons are known, and one more inferred. The most familiar lepton is the electron (usually denoted e). As far as can be told, the electron is a truly structureless particle with no internal parts. It can be envisaged as point-like and elementary. Electrons are all identical, and possess negative electric charge. They are therefore subject to the electromagnetic force.

There are two other charged leptons, the muon ($\mu$) and the tauon ($\tau$). Both are unstable, and decay into electrons after about $10^{-6}$ s and $10^{-13}$ s respectively. In all respects except mass and stability, the muon and the tauon seem to be identical to the electron. The muon's mass is about 200 times that of the electron; a tauon is about 3500 times as massive as an electron. They are simply duplicate 'big brothers'.

The remaining three leptons are neutrinos. These elusive particles carry no electric charge and probably move at the speed of light. Each charged lepton has a neutrino associated with it. Thus there are electron-neutrinos, muon-neutrinos, and tauon-neutrinos, denoted $\nu_e$, $\nu_\mu$, and $\nu_\tau$ respectively. (At the time of writing $\nu_\tau$ has not been identified experimentally.) Most neutrinos interact so feebly with matter that they could easily penetrate light years of solid lead. They have the distinction of being the most common objects in the Universe – there are about a billion neutrinos for every electron or proton – but they all pass straight through us without the least effect.

All six leptons are subject to the weak force and, of course, gravity, but none feels the strong force. This makes their behaviour simpler to analyse, and it has been possible to test the theory of leptons experimentally, which involves subtle quantum and relativistic effects, to an accuracy unprecedented anywhere else in science.

### Quarks

Quarks resemble leptons in appearing to be point-like and structureless, but they differ greatly in their other properties. They all carry electric charge, but in curious fractional quantities of one-third or two-thirds of the fundamental unit of charge carried by the electron. Quarks also feel the weak force and gravity, but in addition they are subject to the strong force, which usually dominates their behaviour.

It is believed that quarks cannot exist in isolation, but only in groups making up composite 'particles' called *hadrons.* Two quark groupings are found in nature. In the first of these, three quarks stick together to make a class of heavy particles called baryons, meaning 'heavy ones'. These include the proton and the neutron. The other arrangement involves just two quarks, making up a class of somewhat lighter particles called mesons. Quarks are much heavier than leptons, and hadrons are hundreds or even thousands of times heavier than the electron.

Just as there are six varieties, or 'flavours' of leptons, so there are six quark flavours (though again, one flavour remains to be verified experimentally). The quark flavours are whimsically given the names up, down, strange, charm, top, and bottom, denoted by u, d, s, c, t, b. Under the action of the weak force a quark can change its flavour, and this can cause the baryon or meson to decay.

Clearly, there are many possible combinations of different quark flavours possible within the hadrons. Each combination is perceived by the experimenter as a different type of particle. Of all these, only the proton is properly stable (and perhaps not even that – see the section 'Unification' below). Those containing s, c, t, or b quarks rapidly decay. During a high-energy collision, dozens of different varieties of short-lived hadrons can be created. Most of them are known only by Greek letters.

## Antimatter

The Theory of Relativity established the fact that mass and energy are equivalent. This means that all forms of energy possess mass and, conversely, that material particles are a form of encapsulated energy. Einstein's famous formula $E = mc^2$ expresses this equivalence.

If enough energy is concentrated, new particles of matter will appear. Thus, the violent collision of two protons, for example, can produce more protons. Whenever matter is created this way in the laboratory it is always accompanied by an equivalent quantity of so-called antimatter. Each lepton and quark possesses a sort of 'mirror', or antiparticle, in which all the physical properties except mass are reversed. Thus, the antielectron (or 'positron') has the same mass as the electron, but opposite (positive) electric charge. Collectively, antiparticles are known as antimatter.

Though antiparticles are often created in laboratory collision experiments, virtually none are observed in nature. If an antiparticle encounters its mirror particle they annihilate each other, releasing energy, usually in the form of gamma radiation. Hence antiparticles tend to meet a rapid demise. Astronomers have used gamma-ray telescopes to place limits on the maximum antimatter content of the Universe, and their conclusion is that the Universe is constituted almost entirely of matter. (The exception could be antineutrinos which, because of their exceedingly weak interaction, can coexist peacefully with neutrinos.) This raises the intriguing question of how the matter of the Universe came to exist in the absence of an equal amount of antimatter.

## The messenger particles

*Photons*

The way in which forces can be transmitted across empty space receives an important new interpretation with the application of the quantum theory. A fundamental hypothesis of quantum physics is that disturbances travelling through the electromagnetic field assume particle-like properties. For example, an electromagnetic wave can be regarded in some respects as a hail of corpuscles called 'photons' in which the energy, $E$, of an individual photon is related to the frequency, $f$, of the wave by the equation.

$$E = hf$$

where $h$ is a universal constant known as Planck's constant. Einstein used the corpuscular nature of light to explain the photoelectric effect in 1905, work for which he was later awarded the Nobel prize. The dual nature of radiation, having both wave-like and particle-like properties, is now a firmly established part of physics.

This 'quantization' extends to the disturbances which travel through the field to bring about the action of one charged particle on another. To take a simple case, suppose two electrons approach each other. Because of their electric charges there will be a repulsive force between the electrons which will cause them to bounce away from each other. Such behaviour is termed 'scattering', and is depicted in figure 1. The electrons in no sense come into direct physical contact. Instead, the force is transmitted as a disturbance through the electromagnetic field surrounding each particle. According to the quantum theory, this propagation of influence through the field does not take place continuously, as implied by the

**Figure 1**
When two charged particles approach, each feels a force of repulsion from the electric field of the other, which causes them to move along diverging paths.

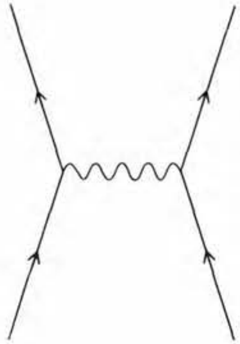


**Figure 2**
A quantum description of the process shown in figure 1 depicts the scattering event taking place abruptly, through the exchange of a 'messenger' photon (wavy line).

smoothly curving trajectories of figure 1, but abruptly, in the form of the exchange of a photon. A more accurate representation of electron scattering is therefore given by figure 2, in which a quantum particle – a photon – passes between the electrons and conveys the force of repulsion. The photon therefore behaves in some sense like a messenger, telling one electron that the other is there and commanding it to move away. The messenger theory can be made mathematically precise, and leads to exceedingly accurate agreement with experiment.

*Gravitons*

Similar theories of the other three forces of nature have been constructed. In the case of gravity the messenger particles, called gravitons, are extraordinarily elusive because of the feeble nature of the gravitational force. For this reason gravitons have not been detected in the laboratory, though their existence is inferred theoretically. The weak force messengers carry the enigmatic names W and Z, and were only identified in the laboratory in 1983. The Z particle is a sort of photon, identical to it in all respects except mass. Whereas the photon is massless, the Z particle is over 90 times as massive as a proton. The W is also extremely massive. It is the high mass of the W and Z that accounts for the very short-range nature of the weak force.

*Gluons*

The strong force is more complicated. It can be considered to arise, in analogy with electromagnetism, by a form of 'charge' carried by the quarks. This charge is unrelated to electric charge, and is given the name 'colour' (it has no connection with actual colour). To account properly for the complexities of the inter-quark force, it is necessary to postulate the existence of three different colour charges – red, green, and blue. The colour is a source of messenger particles just as electric charge is a source of photons. The messengers which transmit the strong force between quarks are called gluons, because they 'glue' the quarks together tightly. Theory demands no less than eight different species of gluons to convey the three different components of colour–charge force.

One curious feature of the inter-quark force is that it does not appear to relinquish its grip on the quarks however hard they are prised apart. The other forces all diminish with distance, but the strong force seems to grow stronger as the quarks move farther apart. This unusual property has led to the hypothesis that isolated quarks cannot exist; only multiple combinations such as the three within baryons are permitted. The hypothesis receives support from high-energy collision experiments, which fail to smash hadrons into their constituent quarks in spite of the huge amounts of energy released.

If quarks are permanently confined within hadrons it leads to an interesting account of the colour concept, according to which 'naked colour' is forbidden. Thus, in a baryon the three quarks always form a red–green–blue, or 'white' combination. In a meson a quark of one particular colour is paired with an anti-quark of the associated anti-colour, thus neutralizing the colour content. Leptons, which are not made of quarks, do not feel the strong force at all; they have no colour 'charge'. The rule 'no naked colour' does not prevent them existing in isolation. We therefore have an elegant explanation for why hadrons are composites whereas leptons come singly.

The quantum description of forces in terms of messenger particle exchange provides a fundamental link between particle physics and forces. Particles act on each other by the exchange of yet other particles. We must place alongside the list of quarks and leptons a third class of particles – the messengers. Evidently the nature of elementary particles cannot be separated from the forces which act on them. The topics of particles and forces are intimately interwoven.

## Unification

A monumental advance in physics occurred when Maxwell unified the forces of electricity and magnetism into a single electromagnetic field of force. More recently, physicists have felt compelled to extend this process of unification. Perhaps all the forces of nature can be combined into just one 'superforce'?

In their search for the superforce, theorists have used certain mathematical guidelines. One of these is symmetry. Symmetry seems to play a fundamental role in the subatomic world, and has already proved indispensable in grouping various species of particle into distinct family units. Two sorts of symmetry are proving valuable in the investigation of unified forces. The first is known as gauge symmetry. A simple example of a gauge symmetry is the freedom to change the zero-point of electrical potential (voltage) without affecting the description of electric forces, which depend only on *changes* in potential. More complicated, but similar, gauge symmetries can be used to describe the other forces of nature. The second type of symmetry, known as super-symmetry, will be discussed shortly.

In the late 1960s Weinberg and Salam discovered a gauge symmetry which simultaneously embodies a description of both the electromagnetic and weak forces, thus providing a unified field theory of the two. Their theory predicted new physical phenomena, such as the existence of the Z particle discovered fifteen years later. The unification process was extended in the mid-seventies with a selection of 'grand unified theories', or GUTs, which incorporate the strong force as well. GUTs are much harder to test experimentally, but they do predict a curious phenomenon – proton decay. The proton, hitherto considered to be absolutely stable, is predicted to have an average lifetime of at least $10^{28}$ years, before decaying into a positron and other particles. The effect is a statistical one (as with all quantum processes), so there is a small chance of one proton among many decaying against the odds during an experimental search. Recent experiments place a limit of $10^{32}$ years on the proton lifetime, so as yet fail to confirm GUTs.

In spite of this experimental uncertainty, physicists have eagerly explored some of the consequences of GUTs. If protons really do decay, the entire Universe will eventually vanish, because the positrons produced by proton decay will annihilate with electrons, leaving only radiation energy. In the decay process hadrons (remember protons are made of quarks) turn into leptons (a positron is a lepton). The two classes of particle, quarks and leptons, were always regarded as distinct. However, GUTs link these two families together just as they link three forces together. They thus provide a unified description of particles as well as of forces. In GUTs new sorts of messenger particles exist which change quarks into leptons, thus allowing protons to decay.

## Creation

Just as GUTs predict that matter will slowly vanish, so they permit the reverse process – the creation of matter from energy. Whereas such a phenomenon is usually accompanied by antimatter, GUTs processes break the matter–antimatter symmetry. Cosmologists believe that they can now explain the origin of matter using these ideas. In the very early Universe, when the temperature was in excess of $10^{28}$ K (this occurred during the first $10^{-35}$ s of the Universe) GUTs processes were greatly speeded up. Heat energy created matter and antimatter, but, because of the GUT imbalance, a slight excess of matter was produced. When the Universe expanded and cooled, the antimatter was annihilated along with most of the matter to produce photons, but a tiny residue of matter remained, eventually to form the Universe we see. A vestige of the annihilation photons still exists in the form of a background of microwave radiation which today bathes the Universe.

The unified theories thus provide a fascinating link between the physics of the very small – the subatomic particles – and cosmology – the physics of the whole Universe. It is a link which is being rapidly developed by theorists, some of whom think that the present structure of the Universe, including the existence of galaxies, was laid down at the earliest moments of existence. They argue that the origin of all fundamental cosmic structures must be sought in the high-energy physics that occurred in the first brief flash that gave birth to the Universe in the so-called Big Bang.

According to this viewpoint, the Universe started in an exceedingly simple structureless state, with no matter and only a single superforce operating. Then, step by step, the different forces of nature separated out and particles appeared, eventually organizing themselves into the forms we now see. The birth of the Universe is thus depicted as a sequence of steps in which more and more structure and complexity 'freezes' out of an initially simple state. Today, we regard the four forces of nature as distinct only because we view them at relatively low energy; likewise the complicated array of subatomic particles we perceive. Just as the simplicity of water freezes into the complex structures of a tangled ice flow, so the ultimate building blocks of matter, exposed only at the ultra-high energies achieved by the Big Bang, are today 'frozen' into the tangle of hadrons and leptons observed by experimenters.

The most ambitious unification schemes are those that seek to combine all four forces of nature. In this respect GUTs only represent an intermediate step because they include only electromagnetism and the weak and strong forces. It is still necessary to incorporate gravity to achieve full unification. One approach makes use of the second type of important symmetry – super-symmetry. This is an abstract concept related to the spins carried by particles, and has no counterpart in classical physics. It provides a possible means of linking all known particles together into a single super-family, including all quarks, leptons, and messengers. Such a theory automatically unifies the forces of nature by placing the messengers of all the forces into a single descriptive scheme. It also unifies force and matter by relating the messenger particles to the others.

These recent attempts at unification, while making rapid progress, are still tentative and fragmentary. Many of the developments are in the form of mathematical models and arguments, and experimental test is often difficult or impossible. Ultimately a truly unified theory of existence will probably have to rest on the force of its philosophical appeal. Meanwhile, new generations of more powerful accelerators are being planned and ingenious experimental techniques devised. The coming years are bound to produce rich new discoveries in this oldest of all scientific enterprises.

## Further reading

CLOSE, F., *The cosmic onion*. Heinemann Educational, 1983.

DAVIES, P. C. W., *The forces of nature*. 2nd edn. Cambridge University Press, 1986.

DAVIES, P. C. W., *Superforce*. Heinemann, 1984.

FRITZSCH, H., *Quarks*. Allen Lane, 1983.

MULVEY, J. H. (ed.), *The nature of matter*. Oxford: Clarendon Press, 1981.

# Radioisotopes

**HESTER GREENSTOCK**
**North London Collegiate School**

The laser was once described as a solution looking for a problem. Radioisotopes have also provided many answers to problems which a few years ago seemed insuperable. Much information on the functions of different parts of the brain has been obtained by attaching electrodes to people's scalps and by testing the behaviour of brain-damaged patients. However, this information cannot be very detailed.

Another technique is now being used in which the subject inhales xenon-133, a $\gamma$-emitter, which then passes into the bloodstream and therefore to the brain. When asked to do simple tasks, like listening to a sound or looking at a simple shape, the areas of the brain which respond receive more blood. The radiation from the xenon in these regions, which is measured by an array of 256 detectors, becomes more intense. Each detector at present monitors an area of $1\,cm^2$, but future techniques will have greater resolution and so many more interesting details about the functioning of the brain as a whole will be observed. For instance, a person reading aloud shows two additional regions, the mouth area and the speech area, to be activated as well as the four areas needed for reading silently. Future developments are likely to provide three-dimensional pictures of the functioning brain. Relatively little is known at present about the regions deep within the brain.

This technique illustrates the ability of radioisotopes to give information about something without disturbing what is being observed; just one of the vast, and increasing, number of ways in which isotopes are being used.

Practically every element has several isotopes. These are atoms with the same chemical behaviour, but slightly different masses. Some isotopes are radioactive and, as it is possible to detect single emissions of $\alpha$-, $\beta$-, or $\gamma$-rays when these decay, minute concentrations can be measured.

Radioisotopes are used in three distinct ways:
as 'tracers' or labels
as clocks
as sources of $\alpha$-, $\beta$-, or $\gamma$-emissions.

## Tracers

The earliest recorded use of a tracer was in 1911 by a young scientist, Georg von Hevesy, who inserted a radioisotope in the leftovers from his dinner. He suspected that his landlady was reheating food from the day before, and the radiation from the next night's dinner proved that he was right! Years later Hevesy was awarded the Nobel prize for his work using tracers in biology.

The chemical identity of a radioisotope with its stable counterpart means that it can be substituted into a particular molecule without changing its chemistry. The radioactive isotope acts as a miniature transmitting station which can send back information, often from inaccessible places like the bottom of oil wells. A biologist can measure the proportion of fertilizer taken up by a crop during different seasons of the year and a geologist can investigate the flow of underground water sources. Entomologists used to track insects by labelling them with coloured dots. Now they can feed them with a radioisotope.

## Labelled compounds

Compounds in which one atom is radioactive are particularly valuable in biological, medical, and biochemical work. Carbon, hydrogen, nitrogen, oxygen, sulphur, phosphorus, and iodine all play important roles in living organisms; unfortunately, however, there are no radioisotopes of oxygen and nitrogen which have long enough half-lives for tracer studies. Carbon-14 and tritium (hydrogen-3) are undoubtedly the two most important isotopes for biological studies since all biologically interesting molecules contain the elements carbon and hydrogen. Sometimes a molecule can be labelled by biological methods of synthesis. A plant or micro-organism is fed on a diet which contains simple molecules labelled with isotopes. The plant or micro-organism then synthesizes complex molecules which have radioisotopes in them. The mechanism of photosynthesis has been deduced by this type of approach.

Many of the reactions investigated are very complicated, but the following example illustrates the technique, even though the 'label' is actually the non radioactive oxygen-18 isotope. Esters are organic molecules which are responsible for the fragrance of fruit and flowers and the 'bouquets' of wines. A typical ester is ethyl acetate, which, under suitable conditions, can be hydrolysed (reacted with water) as follows:

$$CH_3COOC_2H_5 + H_2O \longrightarrow \underset{\text{acetic acid}}{CH_3COOH} + \underset{\text{ethanol}}{C_2H_5OH}$$

The problem to be solved is: where has the water molecule gone to? The answer is obtained by labelling the water with oxygen-18. The reaction will be one of the following:

$$CH_3COOC_2H_5 + H_2{}^{18}O \longrightarrow CH_3CO^{18}OH + C_2H_5OH$$

or

$$CH_3COOC_2H_5 + H_2{}^{18}O \longrightarrow CH_3COOH + C_2H_5{}^{18}OH$$

The mass spectrometer is now used to examine the ethanol formed in the hydrolysis and shows that the oxygen-18 is present in it, rather than in the acetic acid. Perhaps the most spectacular use of the technique has been in breaking the DNA code. This involves discovering what sequence of bases makes the DNA reproduce a particular amino acid, the chemicals from which proteins are built up.

One of the most useful radioisotopes for the agriculturalist is phosphorus-32. Plants need phosphorus as an element of their diet, and fertilizers which contain phosphorus in the form of phosphates are given to them. The total amount of phosphate which a plant takes up during its growth can be determined by chemical analysis, but this will not distinguish between phosphorus from the soil and phosphorus from the added fertilizer. However, using a fertilizer whose phosphate is labelled with phosphorus-32 solves the problem. Now it can be found which phosphate is most effective and also at which stage during its growth the plant's uptake of phosphorus is highest. The farmer can now decide when to fertilize his crops for maximum effect.

**Figure 1**
Labelled fertilizer is being applied to oilseed rape. *Rothamsted Experimental Station.*

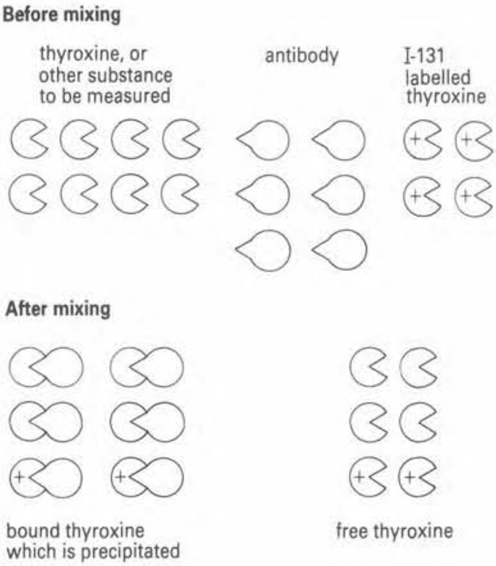
Medical investigations using radioisotopes are either '*in vivo*', which means measuring the radiation from an isotope administered to the patient as a drug, or '*in vitro*', when experiments are done in the laboratory on specimens obtained from the patient. *In vivo* studies can be used to investigate the flow of blood through the heart or venous system, the functioning of the lungs, kidneys, liver, or thyroid gland, and abnormalities in the bone structure, to name but a few examples. Wherever possible the radioisotopes used have short half-lives (a few hours) so that the body is not exposed to radiation long after the measurements have been made. This means that hospitals must have their own generators so that the isotopes to be used are obtained as daughter products from another decay process. It is also preferable to use $\gamma$-emitters rather than $\beta$-emitters. This is because of the greater absorption of $\beta$-particles by the body tissue which both causes radiation damage and decreases the number of particles reaching the detectors. Modern gamma detectors can produce a clear image of the distribution of the isotope with far weaker doses than would be needed if a $\beta$-emitter were used.

We will look at the condition which was the first to be diagnosed and treated with radioisotopes, abnormality of the thyroid gland. Thyroxine is a hormone which is concerned with growth and mental development in childhood and with regulating the rate of metabolism in adults. If a hitherto normal adult has a thyroid gland which makes too little thryoxine that person becomes dull and lethargic and the bodily processes slow down. On the other hand, the thyroid gland may become overactive and produce too much thyroxine. When this occurs bodily mechanisms speed up and the person loses weight and becomes irritable.

Someone with suspected thyroid problems would first be given a variety of biochemical tests, including the measurement of thyroxine in the blood serum. The measurement of very minute concentrations of this and many other hormones, enzymes, and drugs in the blood is possible by the process of radioimmunoassay (figure 2), of which the thyroxine measurement is just one example. A measured volume of the patient's serum is mixed with a known quantity of iodine-131 labelled thyroxine. A known quantity of an antibody to thyroxine is then added. The antibody, in this case, is a protein whose molecules exactly lock into thyroxine. The quantities are such that there is an excess of thyroxine. The antibody combines with the labelled and unlabelled thyroxine in the proportions in which they are present. The combined thyroxine–antibody protein is then separated by precipitation from the excess thyroxine and the activity of the precipitate is measured and compared with standard results. The more thyroxine present in the serum, the smaller the proportion of the labelled thyroxine that will be able to combine with the antibody. This exceedingly sensitive approach can be used to measure minute concentrations (pg $cm^{-3}$) of many hormones and enzymes with no damage to the patient.

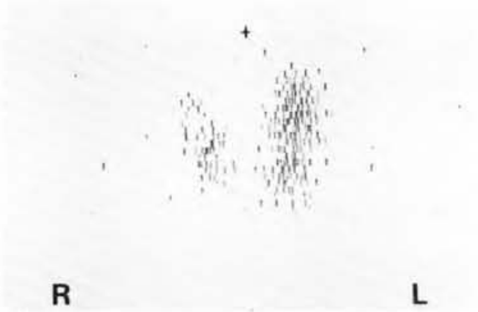
As far as the thyroid is concerned, further investigation can then be carried out *in vivo* if it is thought to be necessary. Thyroxine contains iodine and if a person takes iodine by mouth, the element rapidly accumulates in the thyroid gland, where it is used to make thyroxine. The radioisotope iodine-131, which emits $\beta$- and $\gamma$-rays, is generally used and given as a dose

**Figure 2**
The radioimmunoassay process.




(a)


(b)




(c)

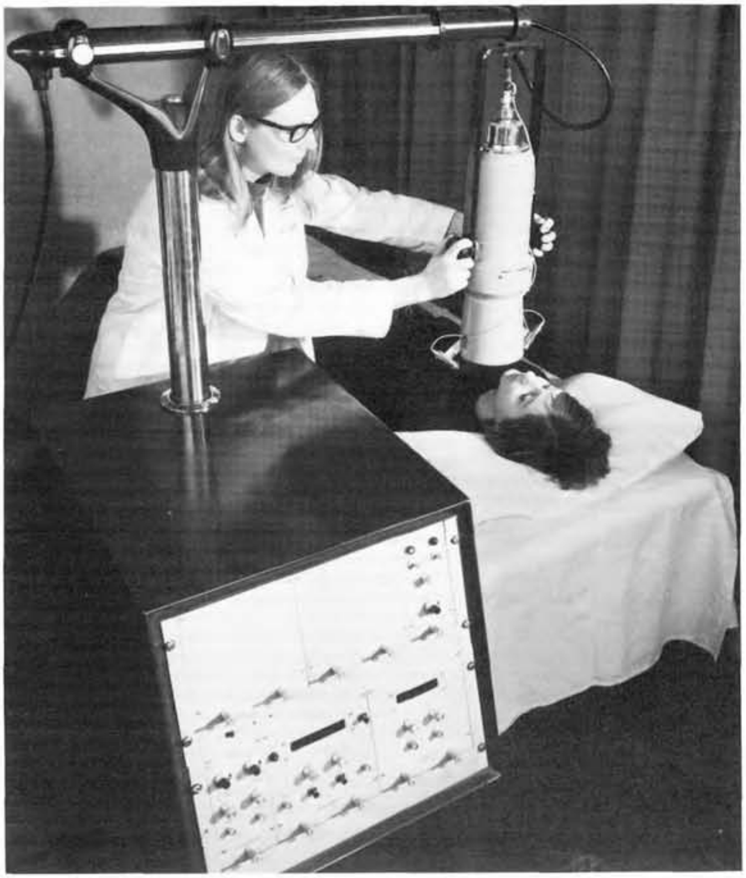
**Figure 3**
Gamma radiation from iodine-131 in three pairs of thyroid glands.
(a) A normal thyroid gland.
(b) One part of the thyroid gland has taken up nearly all the radioactivity. It is therefore overactive, a condition caused by a nodule in the gland.
(c) An underactive thyroid condition.
*Photographs courtesy of the Royal Marsden Hospital.*

of labelled sodium iodide. Some time later the amount of iodine-131 which has accumulated in the thyroid is estimated. This is done by detecting the radiation given off from the gland by means of a gamma detector and comparing the result with those obtained from healthy glands (figure 3). In a case of overactive thyroid a carefully measured dose of iodine-131 can be given actually to destroy a part of the thyroid instead of removing part of it surgically.

## Tracers in industry and the environment

Many processes in industry use tracers as a matter of course. The movement of fluids, their velocities and mixing efficiencies, for instance in a cement mixer, can all be monitored by tracers. Leaks in pipes can be detected by the concentration of activity round the leak. The distribution of industrial effluent, a major cause of pollution, can be studied by measuring the tracer concentrations in areas near the chimneys or outlet pipes. In rivers and estuaries not only the presence of pollution but also the movement of sediment can be detected. The concentration of the undesirable phytoplankton (tiny plant organisms) which grow in reservoirs is measured by exposing a sample of water to carbon dioxide labelled with carbon-14. The greater the quantity of carbon-14 taken up, the greater the quantity of phytoplankton. A similar technique was used to try to detect life on Mars during the Viking probe in 1976.

**Figure 4**
Measuring uptake of iodine-131 by a patient's thyroid gland.
*United Kingdom Atomic Energy Authority.*



## Radioisotopes as clocks

Carbon-14 not only acts as a tracer for detecting small quantities of plant material. It is also well known as a clock. When a plant is living it takes up some of the minute proportion of carbon-14 which occurs naturally in the atmosphere during the process of photosynthesis. The carbon-14 is produced by the action of cosmic rays on the atmosphere. Some nuclear reactions occur which produce neutrons. Then according to the equation:

$$^{14}_{7}N + ^{1}_{0}n \longrightarrow ^{14}_{6}C + ^{1}_{1}H$$

carbon-14 is produced. This isotope is taken up by plants and animals. When they die the $^{14}C$ slowly decays (it emits weak β-particles and has a half-life of 5 730 years), and so the ratio of $^{14}C$ to the total amount of carbon changes. There are obviously a number of assumptions in the measurement.

**1** The proportion of $^{14}C$ in the atmosphere must remain constant.

**2** The mass of $^{14}C$ in all living materials must bear a constant relationship to the total mass of carbon in the material.

**3** The specimen must actually have been growing at the time for which the date is to be measured. A piece of wood from the centre of a tree might be considerably older than the tool into which it was made.

The actual measurement is made harder by the fact that the β-particles emitted have low energies and also that the count rate, even of a young specimen, is considerably lower than the background count.

One way of getting round this problem is to burn the carbon and then introduce it into the counter as carbon dioxide. Another more compact method is to produce benzene from the sample and then to mix it with a solvent–scintillation mixture. With careful elimination of background radiation the maximum age that can be measured is about 40 000 years, but it is hoped to raise this to 100 000 years.

The clock needed to date rocks must have a greater half-life. A variety of decay processes with half-lives of the order of $10^{10}$ years is used. Radioisotopes of the rare earth elements are particularly apt as they occur in suitable concentrations in many types of rock. Potassium-40 (half-life $= 1.26 \times 10^9$ years) decays to produce argon which is trapped in the rock. The relative numbers of atoms of these elements may be measured using a mass spectrometer, and the age then calculated. The age of the Earth, meteorites, and Moonrock have all been estimated by this approach.

## The properties of radiation

### Ionization

The emissions which make tracers so useful cause ionization along their tracks. This ionization is dangerous to living things but also has its direct uses. Static electricity may be a problem in machinery for several reasons. It causes repulsion of paint spray from the surface being painted, discomfort to machine operators, and the possibility of sparks causing fires. The problem can be overcome by using rollers and nozzles coated with polonium. This α-emitter ionizes the air around it and discharges the static charge.

### Biological effects

The fact that radiation has biological effects has been known for a long time. The activity of sources is measured in terms of the number of emissions per second. The unit is the becquerel, equal to one emission per second. For a given sample the activity will depend on the half-life and the number of atoms present. Thus $2 \times 10^{-5}$ kg of strontium-90, of half-life 28 years, has the same activity as $10^3$ kg of uranium, of half-life $4.5 \times 10^9$ years. However, the level of activity alone does not indicate the damage that can be done.

During the course of Marie Curie's work on the separation of radium from pitchblende, Professor Becquerel borrowed from her a specimen of the radioactive material for a demonstration elsewhere in Paris. He put the phial in his waistcoat pocket and was surprised to find the next day that a red patch had developed on his skin beneath the phial. Many of the pioneer workers in the field of radioactivity received severe, and in some cases fatal, doses of radiation. The biological effects are due to the interactions of the highly energetic radiations with the atoms and molecules of the body. Ionization or break-up of the molecules occurs and, if the dose is sufficient, observable effects are produced in the body. Typical symptoms of radiation sickness are dermatitis leading to skin cancers, loss of hair, leukaemia, vomiting, haemorrhages, and diarrhoea. Large doses of radiation will, of course, lead to death. However, radiation concentrated on a cancer can destroy the cancer but not the patient.

Radiation also affects the body in a manner which cannot be seen in the individuals receiving the dose but which appears in their descendants. This genetic effect is due to damage caused to the double helix of DNA, the key genetic molecule which determines an individual's characteristics.

All biological tissue is made up of cells which contain complex molecules, including DNA, suspended in a fluid which is principally water. At birth a baby is made up of about $2 \times 10^{12}$ cells, which increase to about $4 \times 10^{13}$ in an adult. Growth involves processes of cell division, a single cell dividing into two whose genetic characteristics are identical. If a cell is damaged by radiation so that damage occurs to the DNA molecules, then faulty DNA is reproduced as the cell divides. This is much more serious in the developing foetus because of the enormous rate of cell division before birth. At one time it was common practice to X-ray pregnant women to determine, for example, the position in which the foetus was lying in the uterus. Today the same information can be obtained by using harmless ultrasound.

In full grown adults cell division has slowed down but the reproductive organs, the testes and ovaries, need special protection from radiation. For example, the testes produce spermatozoa by processes involving cell division. If radiation damage occurs to cells involved in these processes faulty information will be continually reproduced.

With a sufficiently high dose of radiation the reproductive power of a living organism can be destroyed. This has been used to good effect in the control of some insect pests. A particularly obnoxious insect, the American screw worm fly, was virtually eliminated from the south-western states of America by the use of radiation. The fly lays its eggs in the wounds and sores of farm animals, and the larvae which develop feed on the flesh of their hosts. When they are fully grown they drop off and pupate in the soil. The life cycle takes only four weeks so the pest spreads rapidly. In 1953 the US Department of Agriculture discovered that screw worm pupae irradiated with $\gamma$-rays from cobalt-60 produced sterile male flies. Female screw worm flies, which mate only once, show no preference for fertile mates over sterile ones. Vast numbers of sterile males from $\gamma$-irridated pupae were dropped from aircraft over the island of Curaçao, which was infested with the pest at the time. By 1955 not a single example of infestation was recorded there. The technique is now widely used to control insects over a limited area and it is being developed to deal with widely distributed pests such as the tsetse fly in Africa. Its extermination would save humans from sleeping sickness and greatly increase the size and strength of herds of cattle.

**Figure 5**
A male tsetse fly which, when irradiated, will pass damaged sperm to females in the wild which effectively sterilizes the females. Large-scale releases of sterile males can eventually lead to the collapse of a tsetse population.
*Tsetse Research Laboratory, University of Bristol.*

Figure 6
Variations induced in five types of wheat spike from exposure to X-rays. The process is used in an attempt to produce hybrid varieties of the crop.
*Plant Breeding Institute, Cambridge.*

Low level irradiation of organic material can bring about small and desirable changes, known as mutations, in the genetic material. Many crops developed by normal breeding techniques are produced from a small number of original genes, and are therefore vulnerable to disease. Mutation can produce a gene which increases resistance. The process is also used to produce crop varieties with shorter growing seasons, shorter stems, and larger yields.

## Food preservation

While World food production is being improved by mutation, much of the produce rots before it reaches the consumer. This is a particularly serious loss in hot countries where decay by bacteria occurs so quickly. Irradiation can harmlessly increase the storage life of foods, though the process is complicated by the fact that some physiological process may also be affected, such as the desired ripening of the fruit. The taste is not impaired and the food itself does not become radioactive. At present this method is used in the UK to prepare food for patients who must have sterile diets. It is in general use in several other countries, including the Netherlands and Japan. The technique must be applied with care, but it is becoming increasingly cost effective as the food requirements of the World increase.

Figure 7
Irradiation of mushrooms improves shelf-life by maintaining whiteness and preventing the caps from opening
*Isotron P.L.C., Swindon, Wiltshire.*

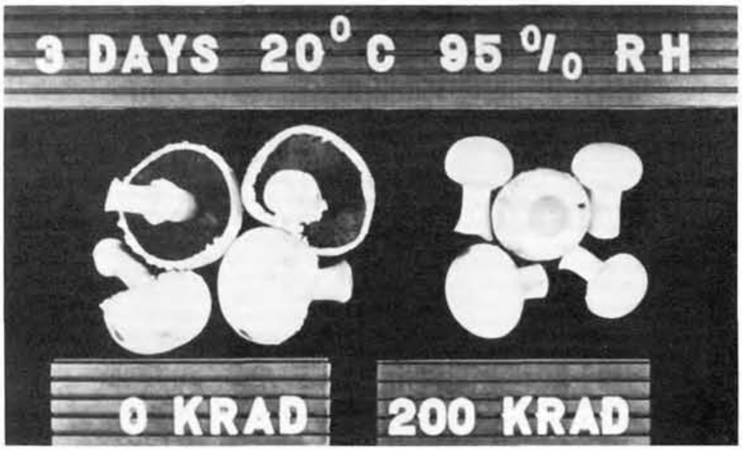
## Sterilization

The complete removal of bacteria can be carried out much more effectively by radiation than by heat treatment. Besides sterilizing food, a considerable industry has grown up to produce sterile food containers, laboratory supplies and, most importantly, hospital supplies. These range from scalpels and dressings to disposable needles and swabs. Also, on an industrial scale, the pathogenic micro-organisms in sewage can be rendered harmless, although this is still generally done by using chlorine.

## Safety

The use of radioisotopes is not without problems, in particular the disposal of radioactive waste. This includes overalls and test-tubes as well as the radioactive materials themselves. Isotopes with relatively short half-lives and high intensity are described as 'hot', and can be isolated until their activity is reduced to more acceptable levels. The waste is then put into sealed containers and either dumped in deep parts of the sea, or in deep, stable, mine-shafts. In either case there is considerable public disquiet as to whether these methods of disposal are 'safe'.

All exposure to radiation produces biological damage, no matter what standards are set regarding minimum exposure. It is always necessary to balance the harmful effects to individuals or their children against the benefits of using radiation at all. We are, of course, all exposed to



'background' radiation in nature. This arises from cosmic rays and from naturally occurring radioisotopes in the atmosphere and rocks. Workers using radioisotopes are given constant checks to make sure that they do not receive excessive doses of radiation. The dose received is usually recorded using a film badge, which is a piece of photographic film in a special holder, attached to the worker's clothing. At suitable intervals the film is developed and the amount of blackening is used to estimate the dose received, which should not be more than 50 times the background radiation dose.

**Figure 8**
Perspex used to shield a researcher from $\beta$-radiation.
*Amersham International, P.L.C.*

Protection by shielding the source from the operator is comparatively simple for β-particles. Even in air, the range of β-particles is limited, being about 1 metre for 0.5 MeV particles and 10 metres for particles of 3 MeV. In dense materials the ranges are much shorter, but the absorption of high-energy particles in matter can give rise to electromagnetic radiation known as *bremsstrahlung* (from the German *bremsen*, to brake, and *strahlung*, radiation). When electrons penetrate matter they are slowed down by electrostatic interactions with the nuclei of atoms. Bremsstrahlung is produced during this slowing down process and is often more penetrating than the β-radiation which produced it. Materials of low nuclear charge produce the least bremsstrahlung and so shielding is made of materials like aluminium, Perspex, and thick rubber (figure 8). Perspex 25 mm thick will absorb particles up to 4 MeV in energy.

Shielding against γ-radiation is more difficult. Unlike charged particles, γ-radiation is not completely absorbed by any material. Instead the intensity of the radiation must be progressively reduced by shielding until it reaches a level safe for those working with it. Concrete, iron, and lead are commonly used as radiation shields. Although potentially dangerous, this penetrating power has many applications in industry.

## Thickness gauges

Radioisotope gauges are used for measuring the thickness of things ranging from electroplated layers and paint films to steel sheet for car body manufacture. Beta particle sources are commonly used for this work, provided that the material is not too thick, say up to 1.5 mm of iron or 4.5 mm of aluminium. For thicker materials γ-sources must be used and these can cope with thicknesses of iron up to 100 mm.

The advantage of radioisotope thickness gauges is that no contact with the material is required and they can be used on continuously produced material which may be passing under the gauge at high speed. This lack of contact is also very valuable if the material is at a high temperature, if it is soft, or if it has a good surface finish which would be scratched by contact gauges.

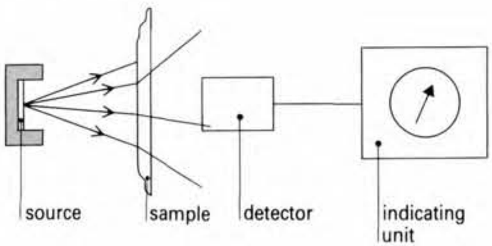



**Figure 9**
Transmission thickness gauge.
*United Kingdom Atomic Energy Authority.*

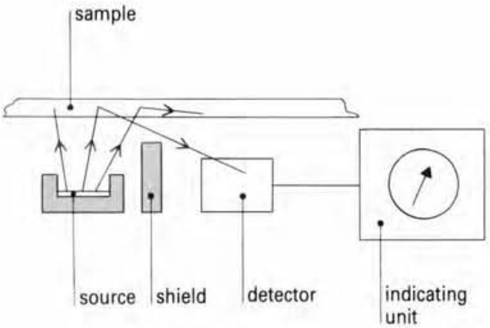



**Figure 10** (left)
Backscattering thickness gauge.
*United Kingdom Atomic Energy Authority.*

**Figure 11** (right)
Radioisotope thickness gauge for measuring thickness of tyre cord using $^{90}Sr$. The gauges are at the top corners of the rubber sheet and measure the amount of radiation which passes through the rubber.
*United Kingdom Atomic Energy Authority.*

Another method, used to measure surface coatings, relies on backscattering of radiation. The amount of backscattering depends on the thickness of the surface coat.

Closely allied to the use of radioisotopes in thickness gauging is the method of checking whether a package has been properly filled. If it is not then sufficient radiation will reach the detector to provide an electrical signal which will activate a device to throw the faulty product off the production line. The same principle can be used to check the level of liquids in tins of food or in sealed storage containers, such as $CO_2$ containers used in fire-fighting.

## Radiography

X-rays have been used for many years as a reliable means of inspecting the interiors of solid objects such as castings, mouldings, and, of course, human beings. Radioisotopes emitting γ-rays are being used increasingly for the radiography of metal objects, the γ-rays replacing the X-rays of the more conventional procedure.

**Figure 12** (right)
Measuring the thickness of a pipe wall to determine the internal area of the pipe using $^{137}Cs$. Gamma-ray backscattering is the principle used.
*United Kingdom Atomic Energy Authority.*

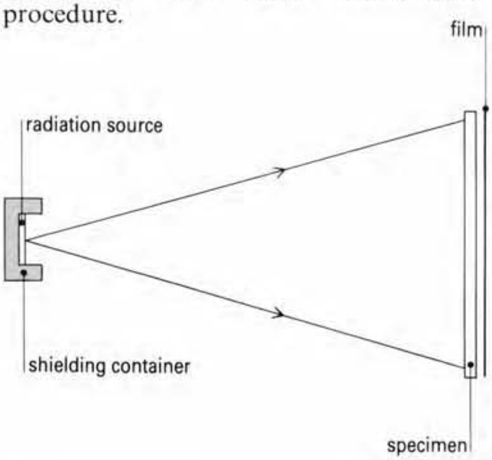



**Figure 13** (left)
Gamma radiation from a radioactive source is passed through the specimen and allowed to fall directly on to a photographic film.
*United Kingdom Atomic Energy Authority.*

A $\gamma$-source of about $10^{12}$ becquerel would be needed to give an output equal to that of an industrial X-ray machine. In practice, weaker sources are used and so longer exposures are needed. The radioisotopes cobalt-60, caesium-137, iridium-192, and thallium-170 are used as industrial $\gamma$-sources.

There are several kinds of situation in which it is more convenient to use radioisotopes than X-rays. Inspection of welds between pipeline sections in the desert is one example. The $\gamma$-source is light and portable and requires no power supplies. Owing to its small size it is often possible to place the isotope source in positions where X-ray sources would not be able to go. Such sources can be used under water to inspect welds in ships' hulls, or the condition of off-shore oil drilling rigs. Corrosion of the legs of these platforms has to be carefully watched. The $\gamma$-radiograph of a pipeline weld is taken by wrapping a photographic filmstrip round the outside of the pipe and then placing the $\gamma$-source in the centre of the pipe for the duration of the exposure.

Radioactive tracers are also used to ensure that oil rig piles are securely cemented to their support sleeves. The foundation cement grout contains a tracer, and so its spread around the pile can be checked.

Beta radiation sources which emit weakly penetrating rays have been used in the radiography of very thin materials, for instance in the detection of water marks in paper.

## Medical use of radiation

Radiotherapy is the branch of medicine in which the destructive power of high energy radiation is used either to destroy or control the growth of unwanted tissue, such as a cancer. This is tissue growing in an uncontrolled fashion. Cancer cells can be destroyed by X- or $\gamma$-radiation, and this much feared disease can often be cured, especially if it is detected in its early stages. Cobalt-60 $\gamma$-sources of up to $10^{13}$ becquerel are used in radiotherapy together with very high energy X-rays. Large

**Figure 14**
A foundation pile for a fixed oil storage/loading buoy in the North Sea. The piles are bonded to their support sleeves by a cement grout monitored by radioisotope tracers.
*Heerema Engineering Service (UK) Ltd.*

**Figure 15**
Arranging the pointer of a powerful radiocobalt unit to ensure that the beam of radiation is accurately directed to a tumour in the patient's lung.
*United Kingdom Atomic Energy Authority. By courtesy of the Hammersmith Hospital, London.*

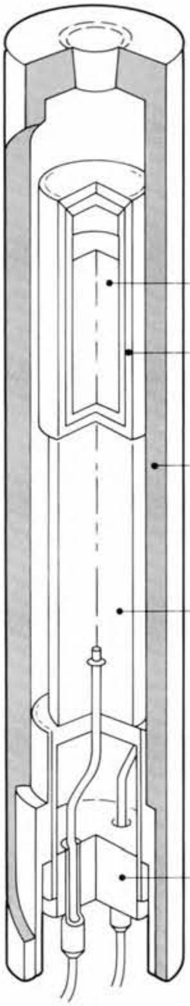



**Figure 16**
Radioisotope powered thermoelectric generator.
*United Kingdom Atomic Energy Authority.*

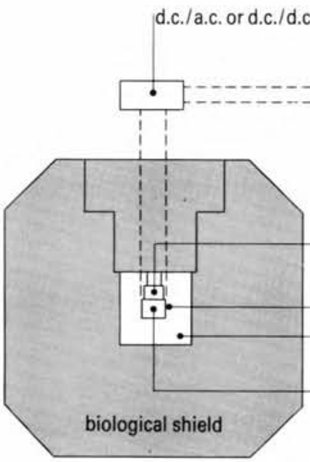



**Figure 17**
Isotope battery (approximate size $45 \times 15$ mm).
*United Kingdom Atomic Energy Authority.*

sources are built up from discs or rods of the radioisotope.

Sometimes the treatment is better carried out by implanting a source within the body for a short period until a suitable dose has been given to the cancer; an example of this is the implanting of iridium-192 wires at the sites of breast cancers.

## Energy from radioisotopes

When the radiation emitted by a radioisotope is absorbed by a material, the material becomes warm. The temperature difference can be used to generate electrical energy. There is an increasing demand for radioisotope-powered generators for space satellites and for terrestrial use in remote places. Such generators, using sources of about $1.5 \times 10^{12}$ becquerel, have been used to power marine navigation lights on buoys, remote weather station installations, submarine cable repeater amplifiers, and land-based navigation lights. They have a life of about ten years, during which they require no maintenance, and this factor helps to offset the cost of the device. About a dozen radioisotopes are suitable as heat sources; strontium-90 and cobalt-60 being two that are often used. The power available from such generators can be hundreds of watts.

It is clear that radioisotopes have their applications in many fields. As new isotopes are developed with emission energies and half-lives of different values their uses will become still more widespread and sophisticated.

## Further reading


AMERSHAM INTERNATIONAL (1983) *Guide for the uses of labelled compounds.* 3rd edn. Amersham International P.L.C.

I.A.E.A. (1976) *Isotopes in day to day life.* International Atomic Energy Agency.

LASSEN, N. A., INGVAR, D. H., and SKINHØJ, E. 'Brain function and blood flow'. *Scientific American*, **239** (4), Oct. 1978, pp. 50–60.

'Researchers set a new date with radiocarbon.' *New Scientist*, **98** (1353), 14 April 1983, p. 79.

# Lasers probe the atomic nucleus

**DEREK EASTHAM**
**Research scientist in the Nuclear Structure Facility division of the SERC's Daresbury Laboratory, near Warrington, Cheshire**

The nucleus sits at the heart of the atom, its constituent protons and neutrons bound together by the strong nuclear force. The energies required to probe the nucleus are therefore high, so it may at first seem surprising that visible light with its moderate energies can reveal important information about the nucleus. Such research has become possible only with the advent of tunable lasers, which can emit light over a range of wavelengths. Experiments with lasers are now being carried out in nuclear physics laboratories throughout the World, revealing clues as to the shape and size of nuclei and providing a precision and sensitivity of measurement not possible before. There are several types of experiment, but I shall concentrate on only two techniques in order to illustrate how lasers come to the aid of the nuclear physicist.

## Energy levels

One very popular experiment is to measure the size and shape of a nucleus and its intrinsic magnetism, or magnetic moment. To do this we look at the very small differences in energy levels of electrons in the atom that these nuclear properties produce. Atomic energy levels depend principally on the nuclear charge, which is determined by the number of protons in the nucleus. The positive charge of the protons generates an electric field, which binds the negative electrons of the atom to the nucleus. If the charged nucleus has a net intrinsic angular momentum, or spin, then the spinning charge acts like a current and produces a magnetic field. This magnetic field influences the energies of the electrons. And if the nucleus is non-spherical the electric field it produces, especially close to the nucleus, alters the atomic energies. These two effects split the atomic energy levels into several components, producing the so-called hyperfine structure.

Another effect that can be measured is the 'isotope shift'. As its name implies, this is a shift in the atomic energy levels from one isotope of a particular element to the next. Isotopes of an element have the same number of protons in the nucleus – and hence the same chemical properties – but different numbers of neutrons, and thus have slightly different masses. Isotope shift is caused partly by the different mass of the nucleus and partly by the alteration in its size. The latter affects the energies because the atomic electrons can actually penetrate into the nucleus, or, in the jargon of quantum mechanics, the wave function describing the position of the nucleus overlaps with that of the electrons. Inside the nucleus, the electric field no longer varies as the inverse-square of the radius, as it does outside the nucleus, and the overlap of wave functions here determines the size of the isotopic shift.

Both the isotope shift and hyperfine splitting are very small in comparison with average atomic transition energies. The differences between electron energy levels in atoms are typically in the region of a few electron-volts, which corresponds to the energy of photons of light in the visible part of the electromagnetic spectrum. The isotope shift and hyperfine splitting might alter these energy transitions by about one part in a million. Nevertheless, physicists have known about these effects for a long time; indeed, measurements of splitting using radio frequency techniques (wavelengths much greater than light with correspondingly low-energy photons) provided the first direct evidence of non-spherical nuclei. The reason for the excitement in using lasers is that they enable us to study unstable nuclei, which was not previously possible. Most known nuclear species are in fact unstable, so the laser provides us with a single technique to measure directly shapes, sizes, and magnetic moments of many more nuclei than we could study before.

### Exotic nuclei

Experiments using lasers give very accurate results; they are also extremely sensitive, so only a small number of atoms is required. This means that we can now study 'exotic' nuclei produced in small numbers in nuclear reactions using particle accelerators. Exotic nuclei have more protons or neutrons than do stable nuclei, and studies of them provide valuable information on the balance of nuclear forces when atoms deviate from the most favourable ratios of neutrons to protons. In addition, the laser methods are so powerful that they can be applied to excited nuclear states. (Nuclei, like atoms, have energy levels; the emission or absorption of a gamma ray photon accompanies the transition between nuclear states.) So far this has been confined to states with lifetimes longer than about one millisecond but experiments are under way to study excited states in nuclei with lifetimes as short as microseconds.

### Resonance fluorescence spectroscopy

There are many different ways to make these measurements on unstable nuclei. I will describe two techniques that employ the method known as *resonance fluorescence spectroscopy* (RFS) for determining atomic transition energies. Both methods are currently being used at the Nuclear Structure Facility (NSF) at the Daresbury Laboratory near Warrington. This is an accelerator constructed for nuclear structure research using heavy ions, and it can produce exotic nuclei.

In RFS a parallel beam of light at a single frequency (or energy) from a tunable laser is directed at the atoms. This light is tuned so that the photon energy exactly matches the desired atomic transition energy. When this happens the atoms absorb photons and spontaneously re-emit light in all directions. The emitted light thus signals the excitation of the atomic transition.

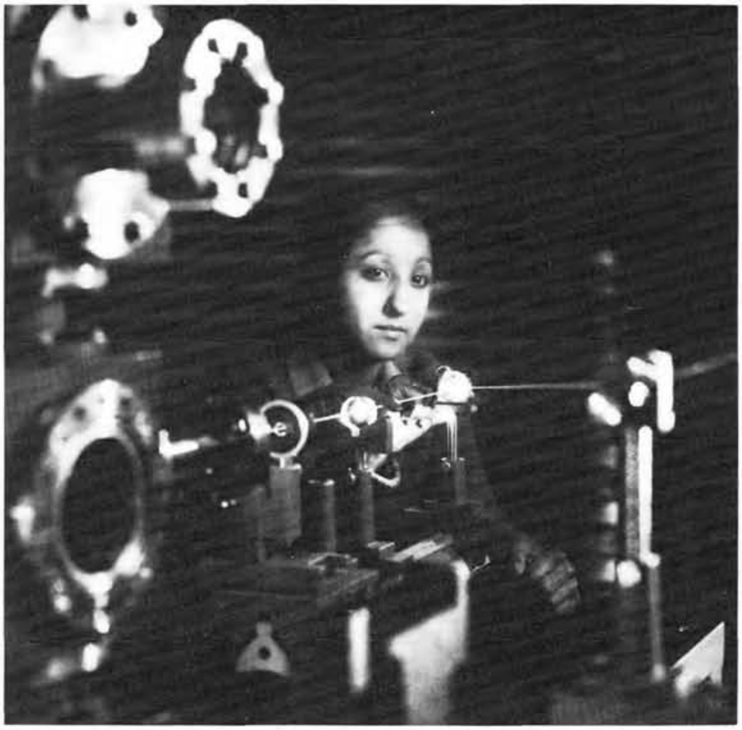


**Figure 1**
Laser light does not interact directly with the atomic nucleus, but it can reveal subtle shifts in the energies of an atom's electrons due to properties of the nucleus. Here a laser beam is being aligned for an experiment to measure nuclear radii.
*Daresbury Laboratory, Science and Engineering Research Council.*

The main problem in these experiments is to overcome Doppler effects caused by the random thermal motion of the atoms. The Doppler effect alters the wavelength and frequency of a wave motion perceived by an observer who is moving relative to the source. It causes the change in tone of the siren as a police car or ambulance passes and the red shift of light reaching us from distant receding galaxies. In RSF experiments the Doppler effect alters the frequency of light 'seen' by the moving atoms. Because the atoms are moving in all directions, they see a broad range of frequencies. So, if a laser is shone directly into a gas the motion of the atoms relative to the laser will broaden considerably the range of frequencies absorbed. To overcome this, the light is shone across (usually at 90°) an atomic beam (figure 2(a), the crossed beam technique), or directly along a beam of fast atoms all moving with the same velocity (figure 2(b), the collinear method).

In the work at Daresbury the atoms to be studied are created in nuclear collisions produced by an accelerated beam of particles from the NSF. The unstable (radioactive) atoms formed in this way are ionized (have one electron removed), accelerated, and separated according to their mass, the separation taking place by means of a magnetic field as in a mass spectrometer. The separated ions are converted back to atoms by passing them through a charge exchange cell containing lithium vapour, from which the ions easily pick up their missing electrons – figure 2(b). The atoms can then be made to interact directly with the laser beam, as in the so-called collinear methods, or they can be converted to an atomic beam by heating in a small, hot tantalum tube – figure 2(a). Both these schemes have their relative merits. A considerable advantage of the collinear technique, however, is the ability to keep the laser frequency fixed while changing the frequency of light seen by the atoms by altering their velocity, using the acceleration voltage of the mass separator.

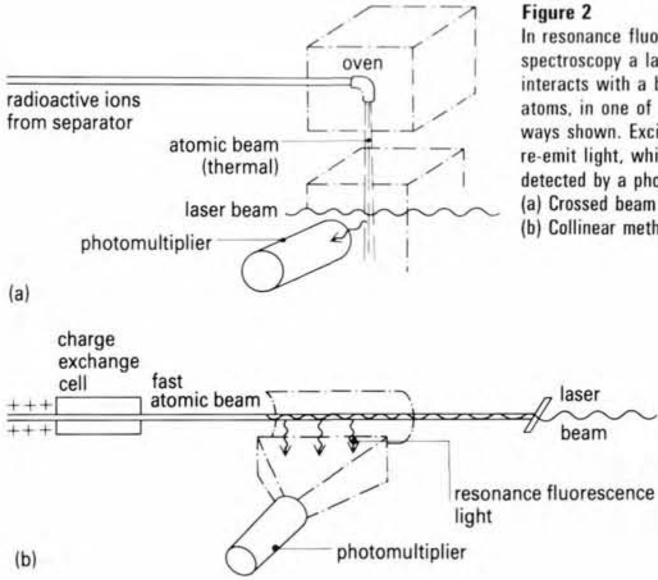



**Figure 2**
In resonance fluorescence spectroscopy a laser beam interacts with a beam of atoms, in one of the two ways shown. Excited atoms re-emit light, which can be detected by a photomultiplier. (a) Crossed beam technique. (b) Collinear method.

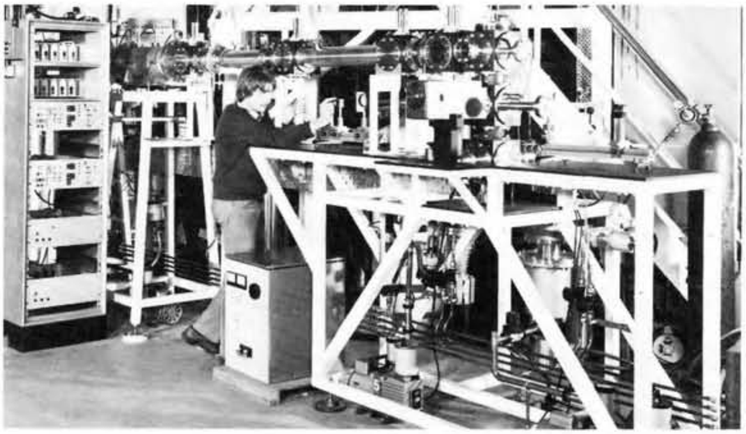
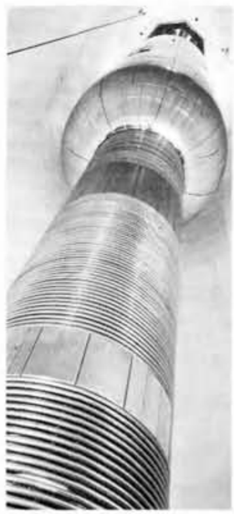


**Figure 3**
A researcher sets up the apparatus to make resonance fluorescence spectroscopy measurements on unstable nuclei. The atomic beam comes in from the left in the horizontal steel vacuum pipe. To the left of the picture is the isotope separator, which uses a magnetic field to separate out different isotopes of the atomic species being studied.
*Daresbury Laboratory, Science and Engineering Research Council.*

(a)



(b)

**Figure 4**
(a) The Nuclear Structure Facility at the Daresbury Laboratory houses a tandem electrostatic accelerator built to produce energetic beams of heavy ions for research in nuclear physics. (b) The vertical accelerator column stands within the tower; the ions are accelerated downwards before being directed to laboratories at ground level.
*Daresbury Laboratory, Science and Engineering Research Council.*

Figure 5 shows the results of some work using an atomic beam containing all the stable isotopes of the element samarium (atomic number, or number of protons, 62). Each isotope is clearly resolved. This would not be the case if a gas cell were used, as the Doppler broadening would cause the peaks to merge together. From these results the change in the nuclear charge radius with isotope can be obtained (figure 6). (The data shown here come from the work of John Griffith and colleagues at Birmingham University who have pioneered many of the techniques in this field.) The nuclear size changes in a fairly smooth way above and below nucleon number 150 (total number of protons and neutrons) where there is a sudden change in average radius. This change shows up directly in the uneven spacing of figure 6, and corresponds to a change in shape of the nucleus from a sphere to a prolate spheroid, or 'rugby ball' shape.

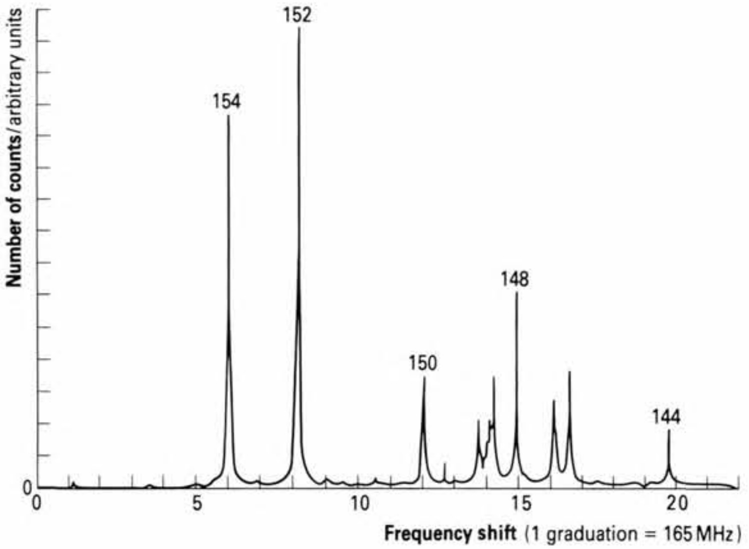



**Figure 5**
Resonance fluorescence spectra from a beam of samarium atoms, showing clear peaks from the various isotopes. The numbers give the total number of protons and neutrons (the nucleon number).

## Models of the nucleus

Changes in nuclear shape are an extremely sensitive way to test our ideas concerning the nucleus. These ideas are usually framed in terms of 'models' because it is not possible to solve exactly the complex quantum mechanical equations that determine the way in which neutrons and protons are bound together in a nucleus.

One simple model of the nucleus, known as the liquid-drop model, likens the nucleus to a simple electrically-charged drop. Of course, the parameters that determine the drop's energy refer to nuclear matter and not to any familiar liquid. This model allows us to calculate the average trends in the binding energies of nuclei, but it predicts incorrectly that all nuclei should have a spherical shape, with a radius which varies as the cube root of its mass, just as the radius of a liquid drop varies with the cube root of its mass. To make predictions about nuclear shapes a model needs to include aspects of the individual motion of the neutrons and protons.

The nuclear shell model does just this and assumes that the individual neutrons and protons move in an average spherical

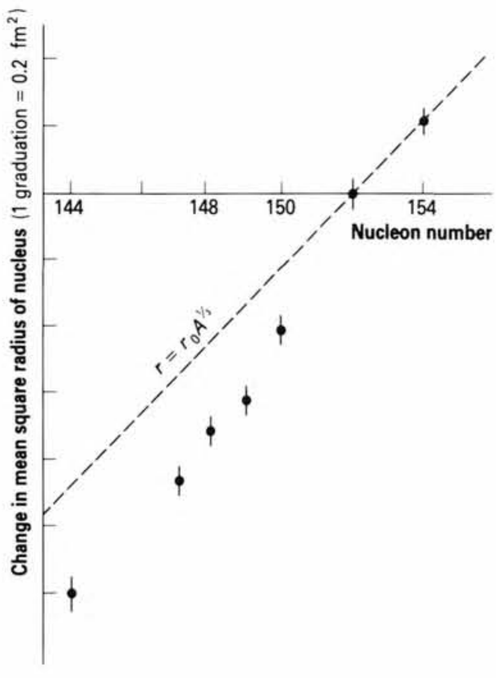



**Figure 6**
The frequency shifts in the spectra give a measure of the nuclear charge radius. The uneven spacing corresponds to a change in shape from a sphere to a 'rugby ball'.
[1 fm (femtometre) $= 1 \times 10^{-15}$ m.]

or deformed (spheroidal) potential field. The shell model can be combined with the collective aspects of the drop model to produce formalisms which are called macroscopic–microscopic models. Although these have had a great deal of success in calculations of nuclear masses and shapes, recent laser experiments have produced some results that do not agree with them. For example, the new work shows that in some regions of the Periodic Table the average change in radius along an isotopic chain (that is, for the isotopes of one element) does not vary with the cube root of the mass. This can be understood by assuming that the neutrons do not mix uniformly with the protons but tend to form a skin on the surface of the nucleus. Indeed it is possible to construct a model, called the droplet model, that puts neutrons and protons in separate concentric liquid drops of different radii. Such a model facilitates accurate predictions of the average change in charge radii for nuclei. The next stage might be to combine the droplet model with the shell model, though no-one has done this yet.

Another aspect of the data that has not been adequately explained is known as odd–even staggering. The improved accuracy of the laser measurements reveals that along an isotopic chain the radius of the nuclear charge does not vary smoothly but shows a saw-tooth effect, with a nucleus with odd mass number often lying below the straight line between the two adjacent nuclei with even mass number. This illustrates dramatically the influence of the individual nucleons on the nuclear size, but as yet no quantitative calculations have been able to reproduce the effect.

## Isotope detection

A second type of experiment at the NSF concerns the use of lasers for ultrasensitive detection of isotopes and elements. Figure 5 shows that as well as providing values of isotope shifts we can use RFS to measure the amount, or abundance, of a particular isotope in a sample. Experiments that can accurately measure the abundance of naturally-occurring rare isotopes, such as beryllium-10, carbon-14, aluminium-26, silicon-32, chlorine-36, and iodine-129, would be important in archaeology and geology, because the proportion of these isotopes can give a measure of the ages of rocks and artefacts. Normal methods of mass spectroscopy do not work in these cases because the isotopes are often obscured by molecular impurities of the same mass. With laser techniques, however, the molecular energy levels of the impurities are quite different from the atomic ones of the rare isotopes.

A most important consideration in laser techniques is the sensitivity of detection, which depends on the probability (the 'cross-section') for absorption of photons by atoms. Because this is a resonance process the effective area or cross-section of each atom is increased by perhaps 100 000 times. This means that certain atoms, which decay straight back to the ground state, can be made to absorb and re-emit thousands of photons during their transit across the laser beam. The detection of single atoms in this way has caused great excitement because of its application in many areas.

One of the problems with laser techniques, however, is that often the first excited atomic transition – to the next available energy level of the atom – corresponds to energies in the ultra-violet part of the spectrum, which is as yet inaccessible to lasers. We can overcome this by employing the simultaneous absorption of two photons or by using a molecular compound containing the element of interest, although both these methods have certain drawbacks.

## Resonance ionization spectroscopy

Another similar technique to RFS is known as resonance ionization spectroscopy (RIS). This was pioneered by scientists at the Oak Ridge National Laboratory in the US specifically for detecting single atoms of one particular element. In this process electrons are pumped into the first excited atomic level by absorbing light from a laser tuned to the correct frequency. But instead of allowing the excited atom to decay by spontaneous emission, as happens in RFS, the atom is photoionized using a second laser beam. (When the ionization potential is large, processes of three or more steps are required, involving other atomic levels.) The advantage of this scheme over RFS is that ions are detected rather than photons and there are no background problems from scattered photons. Experimentally the system is extremely simple and consists of a device rather like a Geiger counter filled with an inert gas and containing the atoms of the element in question. The laser beam(s) are shone directly into the gas and ions are detected as pulses across the high-voltage electrodes.

One of the many proposals for using RIS in nuclear physics concerns the mass of the neutrino and its implication in the missing mass of the Universe. The neutrino (symbol $\bar{\nu}$) is produced in the beta decay of nuclei, when a neutron converts into a proton and an electron, for example, in the decay of protactinium:

$$^{234}_{91}\mathrm{Pa} \longrightarrow {}^{234}_{92}\mathrm{U} + {}^{0}_{-1}\mathrm{e} + \bar{\nu}$$

Theories of the electroweak interaction which describe this process also predict that it is possible for the nucleus to decay by the simultaneous emission of two electrons or positrons.

This process is known as double beta decay. An example is:

$$^{130}_{52}\mathrm{Te} \longrightarrow {}^{130}_{54}\mathrm{Xe} + {}^{0}_{-1}\mathrm{e} + {}^{0}_{-1}\mathrm{e} + 2\bar{\nu}$$

The rate for double beta decay is extremely sensitive to the mass of the neutrino and would provide an accurate method for measuring it. However, the half-life of the process is something like $10^{22}$ years; in other words, half a sample of nuclei of tellurium-130 would decay by this means in $10^{22}$ years. The technique therefore requires extremely sensitive methods. One proposal is to take a 2 kg piece of tellurium and leave this for a year; tellurium is chosen because it transmutes into xenon, which is unreactive and relatively easy to detect. During this time approximately 1000 of the $10^{25}$ atoms of tellurium in the sample would change to xenon by double beta decay, and the xenon atoms could be detected by RIS methods.

## Proton decay

Another important rare process that lasers may help us to detect is the decay of the proton. The proton is very stable and its lifetime in a nucleus is predicted to be greater than $10^{28}$ years, so it is not feasible to wait for a sufficient number of transmuted atoms to accumulate. But it is possible to assess the ratio of elements in geological samples, which have been present since the formation of the Earth some four thousand million years ago. This geological technique may also be used to estimate the flux of neutrinos from the Sun during the Earth's history.

Although I have described two applications of lasers in nuclear physics there are many others, and the number continues to increase as laser technology advances. The ideal laser for nuclear physics would be one working in the gamma ray part of the spectrum, called a GRASER. Such photons could be made to interact directly with the nucleus, where transition energies are comparable with gamma ray energies. The problems in constructing such a device seem at present insurmountable, but who can predict the future?

(This article first appeared in *New Scientist* **97**, 1350, March 1983, 789–792. It has been amended slightly by the author.)

# Imaging

**CHARLES TAYLOR**
**Emeritus Professor of Physics at University College, Cardiff, and Professor of Experimental Physics at the Royal Institution, London**

## 1 Introduction

Images play an important part in modern life: holiday snaps, cinema films, telescopes, microscopes, television, navigation by radar or sonar, and being scanned by ultrasonics or X-rayed at the hospital are examples. On the face of it they seem quite unconnected, but, in fact, all involve waves of some kind and it is my aim to show how they all have physical principles in common.

## 2 The two stages in the imaging process

The cameras in a television studio are useless unless the scene is properly lit with light waves; radar needs radio waves, sonar needs sound waves and, whatever the radiation used, the first step in any imaging process is the interaction between waves and an object. During this interaction the waves are changed in some way and so are able to carry with them some information about the object. But without some kind of focusing or scanning system we cannot interpret this information.

My favourite illustration of this (which you can easily try for yourself) is to put a slide in an ordinary slide projector but with the lens removed. All that you can see on the screen is a fuzzy patch of white light; but the information about the slide must be there on the screen in some form because the lens cannot add anything when it is replaced. All it can do is to sort out, or decode, what is already there. Put crudely, the lens does not *know* what is on the slide, but it does know how to sort out the information to give an image.

So, the essential stages are:

1 Scattering of the waves by the object.
2 Recombination of the scattered waves to form the image.

If an object happens to be self-luminous, as for example a street lamp or a star, the first stage is not scattering, but radiation; the information is 'ready-coded' in the waves as they leave, but thereafter the sequence is exactly the same as for non-luminous objects.

## 3 How are the details of the object carried by the waves?

It is easy to see how some details are carried. For example, if waves of white light interact with a red flower, the shorter wavelengths (blue, green, etc.) are absorbed and only the longer wavelengths carrying the information that the flower is red are scattered. But what about the surface texture, the size and shape, and so on? How are they encoded? If it was by some change in *wavelength*, then, even with a black and white slide, the fuzzy patch of light in the lensless projector experiment would be coloured – and it isn't. If the object changed the *amplitude* of the wave to encode this information then we would see light and dark areas in the fuzzy patch – and we don't see this either. The only other parameter of the wave that we can change is its *phase*. But how could phase changes be used to convey all the detailed information involved in an image?

In figure 1(a) we will imagine that the points A and B are points on the slide in the lensless projector experiment; the black and white striped lines represent the crests and troughs of waves travelling to point C in the fuzzy patch on the screen; you can see that they arrive in phase with each other. (A working model using striped string to represent the waves is more effective than the diagram.) In figure 1(b) you can see that at two other points on the screen, D and E, there are phase differences between the waves from A and B. In figure 1(c) the points A and B have been moved closer together; the waves are still in phase at C. But in figure 1(d) you can see that the phase differences at D and E are less, and even at E the phase difference is still only as much as it was at D in figure 1(b). I hope this makes clear that we are encoding information about the relative positions of A and B in the phase of the waves. The whole object could be regarded as made up of pairs of points such as A and B, and hence, in principle, all the object details could be encoded.

But what happens if the points A and B are moved even closer together? Figure 1(e) shows that the waves are still in phase at C, but at D and E in figure 1(f) there is still very little phase difference; even as far round as F there is still less than 90° difference. The important principle thus emerges that if the points are much closer together than about one wavelength no information about them can be encoded.

It follows that, if we cannot encode much information, then no matter how powerful the rest of our system is we shall not be able to form an image. The lower limit of detail that can be imaged is about one wavelength, though we shall say more about this in Section 7. Table 1 shows the wavelengths of some waves that are used in imaging together with the sizes of common objects that one might wish to observe.

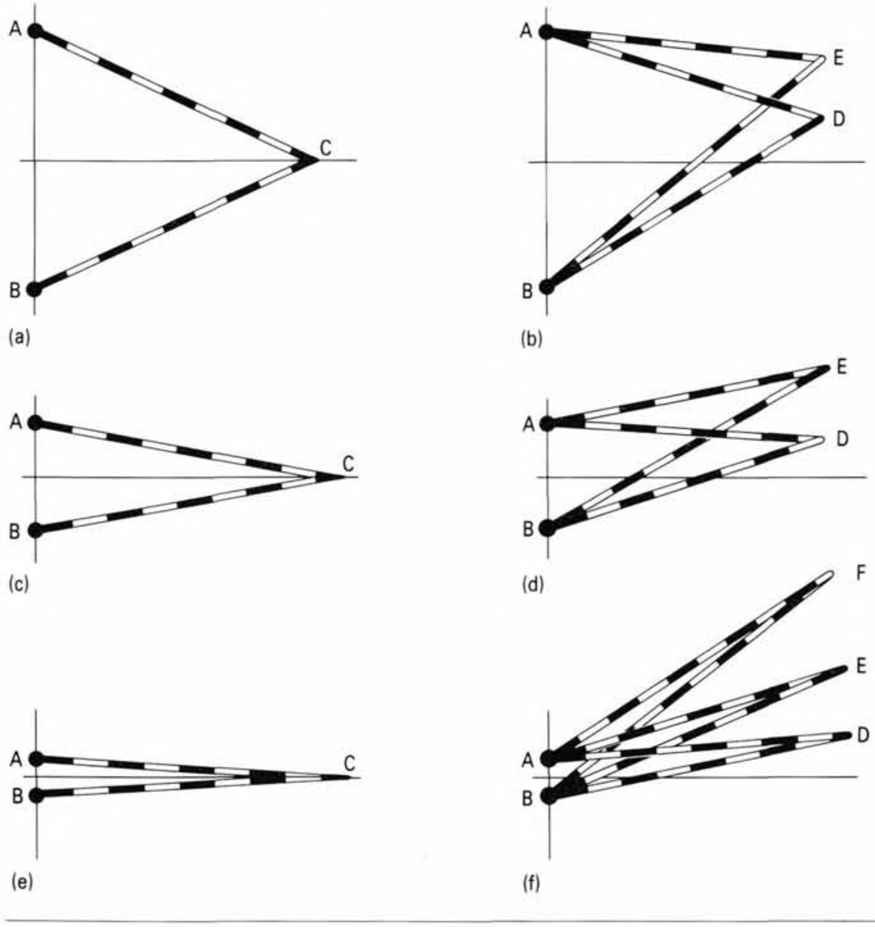



**Figure 1**
The black and white stripes represent crests and troughs of waves scattered from two points, A and B, on the slide in the lensless projector experiment.
(a) Waves arrive at the screen (in the fuzzy patch of light) at C in phase.
(b) Waves arrive at D with a phase difference of 180° and at E the phase difference is more than 360°.
(c) The points A and B are now closer together but the waves are still in phase at C.
(d) Now the phase change at D is quite small and we have to move to E to achieve a phase difference of 180°.
(e) The points A and B are now even closer, they are less than a wavelength apart but the scattered waves are still in phase at C.
(f) The phase differences at D and E are now very small, and even right round at F the phase difference is barely 90°.

| Typical objects in the size range | Order of magnitude in metres | Frequency in hertz of sound waves having this wavelength | | Frequency in hertz of electromagnetic waves having this wavelength | |
|---|---|---|---|---|---|
| City | $10^4$ | | | $3 \times 10^4$ | |
| Airfield | $10^3$ | | | $3 \times 10^5$ | |
| Jumbo jet | $10^2$ | 3.3 | Infra sound | $3 \times 10^6$ | |
| House | 10 | $3.3 \times 10$ | | $3 \times 10^7$ | Radio |
| Child | 1 | $3.3 \times 10^2$ | Audio sound | $3 \times 10^8$ | |
| Sparrow | $10^{-1}$ | $3.3 \times 10^3$ | | $3 \times 10^9$ | Infra-red |
| Bee | $10^{-2}$ | $3.3 \times 10^4$ | Ultrasonics | $3 \times 10^{10}$ | |
| Daphnia | $10^{-3}$ | $3.3 \times 10^5$ | | $3 \times 10^{11}$ | |
| Amoeba | $10^{-4}$ | $3.3 \times 10^6$ | | $3 \times 10^{12}$ | |
| Chlorella | $10^{-5}$ | $3.3 \times 10^7$ | Acoustic surface waves | $3 \times 10^{13}$ | Visible light |
| Typhoid bacillus | $10^{-6}$ | $3.3 \times 10^8$ | | $3 \times 10^{14}$ | Ultra-violet |
| Smallpox virus | $10^{-7}$ | $3.3 \times 10^9$ | | $3 \times 10^{15}$ | |
| Turnip yellow virus | $10^{-8}$ | $3.3 \times 10^{10}$ | | $3 \times 10^{16}$ | |
| Vitamin A molecule | $10^{-9}$ | | | $3 \times 10^{17}$ | X-rays |
| Carbon atom | $10^{-10}$ | | | $3 \times 10^{18}$ | |
| | $10^{-11}$ | | | $3 \times 10^{19}$ | γ-rays |
| | $10^{-12}$ | | | $3 \times 10^{20}$ | |
| | $10^{-13}$ | | | $3 \times 10^{21}$ | |
| Uranium nucleus | $10^{-14}$ | | | $3 \times 10^{22}$ | |

**Table 1**
This shows the relationship between the size of objects that might be imaged and the wavelengths and frequencies of different kinds of radiation.

## 4 Why do we not see detail in the scattered light?

When two beams of light arrive at the same point simultaneously the phenomenon of interference may occur and the resultant pattern will depend on their relative phase. Indeed, figure 1 is very similar to the kind of diagram that illustrates the origin of Young's interference fringes. Why then do we not see a complicated fringe pattern in the fuzzy patch of the lensless projector?

To understand the reasons for this we need to introduce another property of waves called 'coherence'. A complete understanding needs some rather complicated mathematics, but an idea of its meaning can be gained by some imaginary experiments with a ripple tank. A single dipper, driven up and down regularly, produces waves like those in figure 2(a). Corks floating on the surface at p and q would move up and down in step with each other and we could say that there is lateral, or spatial, coherence between them. Several irregularly spaced dippers, all moving up and down regularly would produce a pattern like that in figure 2(b) and, though prediction of exactly how the two corks would move would be quite difficult, it is clear that they would no longer move in step; we could say that they are spatially incoherent. A small slit placed in the tank as in figure 2(c) restores the spatial coherence and this is why we use a source slit in experiments such as Young's fringes.

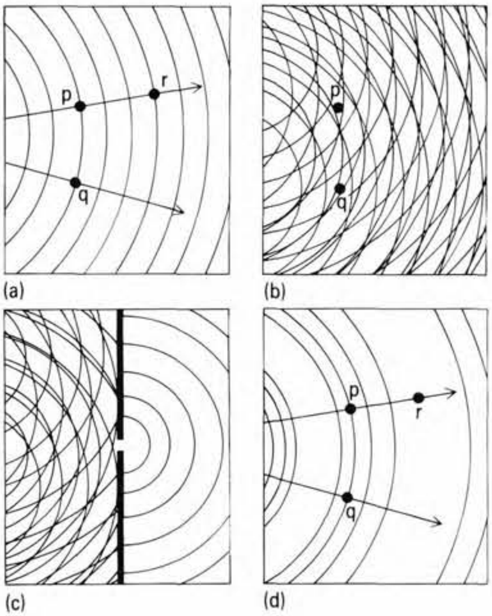



**Figure 2**
The lines represent crests of waves on the surface of water in a ripple tank.
(a) Regular waves from a single dipper; corks floating at p and q will move up and down in step, showing spatial, or lateral coherence; corks at p and r will also move up and down in step, showing temporal, or time coherence. (Of course the waves at p and r are 3 × 360° out of step, but the effect is the same as if they were in phase.)
(b) Regular waves from many irregularly spaced dippers; there is now no simple relationship between the corks at p and q and they show lateral *in*coherence.
(c) Lateral coherence can be restored by the introduction of a slit.
(d) Irregular waves from a single dipper; corks at p and q still perform the same movements and so we still have lateral coherence; corks at p and r would no longer be moving up and down at the same frequency at a given instant and could be said to display temporal *in*coherence.

But there is another kind of coherence called temporal, or time coherence. In figure 2(a) a cork at r would move up and down exactly as does p, although in fact the waves do not reach r until a short while after leaving p. Figure 2(d) shows what happens if the single dipper is *not* vibrating regularly; p and q still move in step and so there is still spatial coherence, but r will no longer be moving in step with p. Indeed, with irregular frequencies like this it would no longer be meaningful to talk of a phase difference between p and r; at a particular instant they may be vibrating at different frequencies. Thus in figure 2(d) we have spatial coherence but temporal *in*coherence. In interference experiments with light we use monochromatic (single frequency) light to ensure temporal coherence.

Now, in the lensless projector experiment the light comes from a large source rather than from a slit and is thus spatially incoherent; also it is 'white', that is, it consists of short bursts of waves (photons) of many different frequencies, and is thus temporally incoherent. Each wave packet will produce its own fringe pattern momentarily on the screen but the whole will change so rapidly that we see merely the fuzzy patch.

The gas laser has revolutionized experimental optics because it produces light that is both temporally and spatially coherent without the need for slits or colour filters and, in addition, is of very high intensity. We can repeat the lensless projector experiment with a laser beam. A piece of gauze, F, placed in the beam can be imaged on a screen S with a lens C (figure 3). But, if the lens is removed we now have a distinct pattern in place of the fuzzy patch. Figure 4(a) is the pattern and (c) is the image of the gauze: the experiment also works with a more irregular object. Figure 4(b) is the patch when a piece of stretch nylon (as for tights) is used, and figure 4(d) is the image.

The fuzzy patch in the lensless projector experiment and the pattern on the screen when a piece of fabric is placed in the laser beam can both be described as 'holograms'. The word simply means 'whole writing', and describes the fact that waves from any *one* point on the object reach *all* points within the patch and, conversely, any *one* point on the screen receives light from *all* points on the object. (See figure 5.)

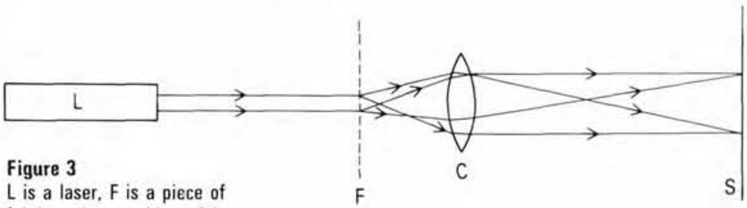



**Figure 3**
L is a laser, F is a piece of fabric acting as object, C is a converging lens which recombines the scattered light to form an image of the fabric on the screen S.

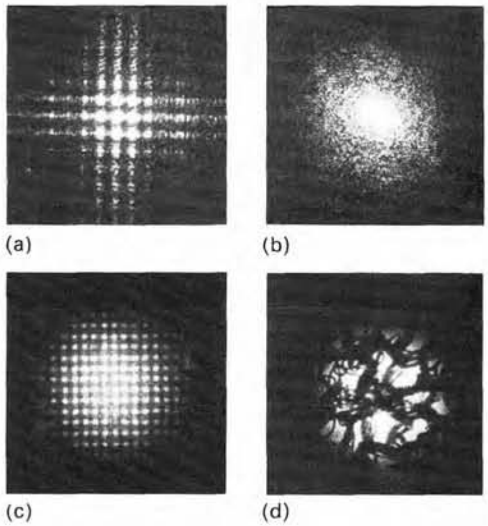



**Figure 4**
If the lens C in figure 3 were removed patterns equivalent to the fuzzy patch in the lensless projector experiment would appear on the screen.
(a) Regular gauze.
(b) Irregular fabric (knitted stretch nylon as for tights).
With the lens in position the image of the object is seen on the screen.
(c) Recombined image from (a).
(d) Recombined image from (b).
*Professor C. A. Taylor.*

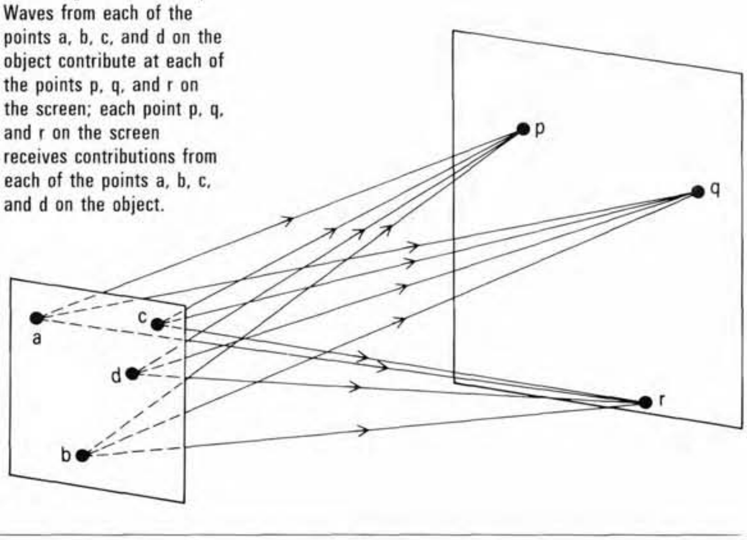



**Figure 5**
The hologram relationship. Waves from each of the points a, b, c, and d on the object contribute at each of the points p, q, and r on the screen; each point p, q, and r on the screen receives contributions from each of the points a, b, c, and d on the object.

You can demonstrate the truth of this principle by using a simple slide (for instance a black cross) in the lensless projector. An auxiliary, long-focus, converging lens can be used to form a small image of the cross on the screen and wherever the lens is placed within the projector beam a complete image of the cross will be produced. Thus *every* point in the patch on the screen contains the whole information about the cross. (If the projector is about 4 metres from the screen, a + 1D lens placed a little over a metre from the screen should be about right.)

## 5 Decoding the information to form an image

In a pin-hole camera we merely throw away most of the energy and the resulting image is rather dim. Figure 6 shows, in two dimensions, how points a, b, and c are imaged at p, q, and r.

Several pin-holes would each produce an image, but at different places on the screen. A small prism of suitable angle could be adjusted over each hole to deviate the images so that they coincide to produce a single, brighter image. If we now imagine increasing the number of holes and prisms until the holes fuse into a single large aperture, we shall find that the prisms will fuse into a converging lens and so will give us a much brighter image. Figure 7 shows how the transition from prisms to lens occurs.

Thus the lens decodes without wasting energy; but there is a penalty. A pin-hole produces an image of similar clarity at any distance; but, just as the prisms will need to be adjusted differently (and different prism angles chosen) for different screen and object distances, so a lens will only produce a sharp image at specific distances. We have met the need to 'focus' the image. But have you ever stopped to think exactly what you do when you focus an image? If the object is a black and white drawing with clean edges to the lines we merely make the picture 'sharp'; if the object is a fuzzy one it may be more difficult and we may find ourselves focusing on dust or small hairs on the slide and hoping that the rest is then in focus.

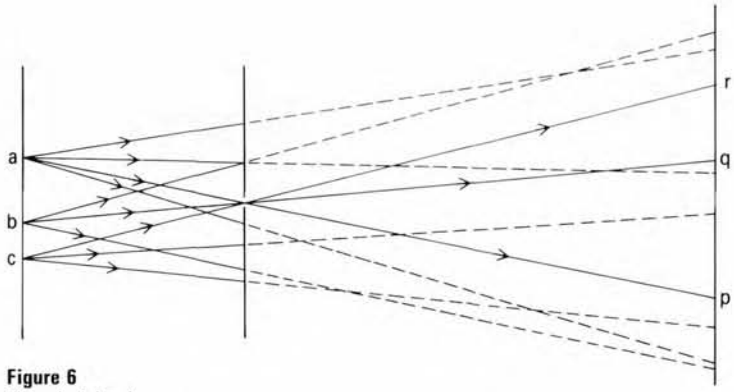



**Figure 6**
The principle of recombination to form an image using a pinhole; most of the light is thrown away.

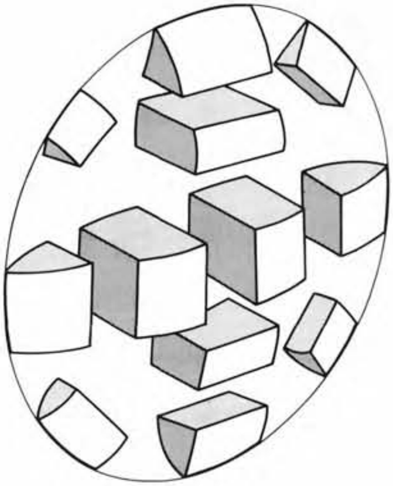


**Figure 7**
A converging lens seen as a collection of prisms.

In fact, what we are doing is to make the object look like we think it *should* look. There is only one other way of focusing, and that is to determine all the focal lengths with great precision and then to adjust all the distances to fit the calculation; in practice this is rarely possible. What we are really saying is that you can only focus properly if you have a pretty good idea of what the object looks like in the first place. In the case of a slide projector, or of a telescope viewing ships or birds, there is no problem. But, if you were the first person to discover a particular object under a microscope you could alter its appearance by changing the focus and have no way of telling which is the 'correct' image. In the nineteenth century when new microscopic objects were being discovered almost daily there were often arguments about the exact appearance of a tiny organism for this very reason.

Lenses work well for light, and electromagnetic lenses can be used to decode the information from scattered electrons in the electron microscope, but sound waves or radio waves are more difficult to focus. Suppose we think again about the pin-hole camera and move the pin-hole back until it is in contact with the object. Now we are cutting out the light from most of the object and allowing the waves from just one particular spot to reach the screen. A detector of some sort (for light a photo-cell could be used, for radio a receiver, and so on) placed anywhere within the beam would then record the information about that one spot on the object. The pin-hole could then be moved to each point of the object in turn and so the complete information could be built up.

We have converted the two-dimensional information that a lens could decode all at once into a series of separate bits of information that follow each other in time. The process is called 'scanning', and it is now probably the most powerful of the imaging techniques available. Its special importance is that it not only solves the lens problem for waves that are not so easy to focus as light, but it also converts the complicated information in a two-dimensional picture into a string of separate bits that can be transmitted over a single telephone line or stored sequentially in a computer memory. Figure 8 shows how a simple picture can be converted by scanning into a linear sequence and subsequently recombined.

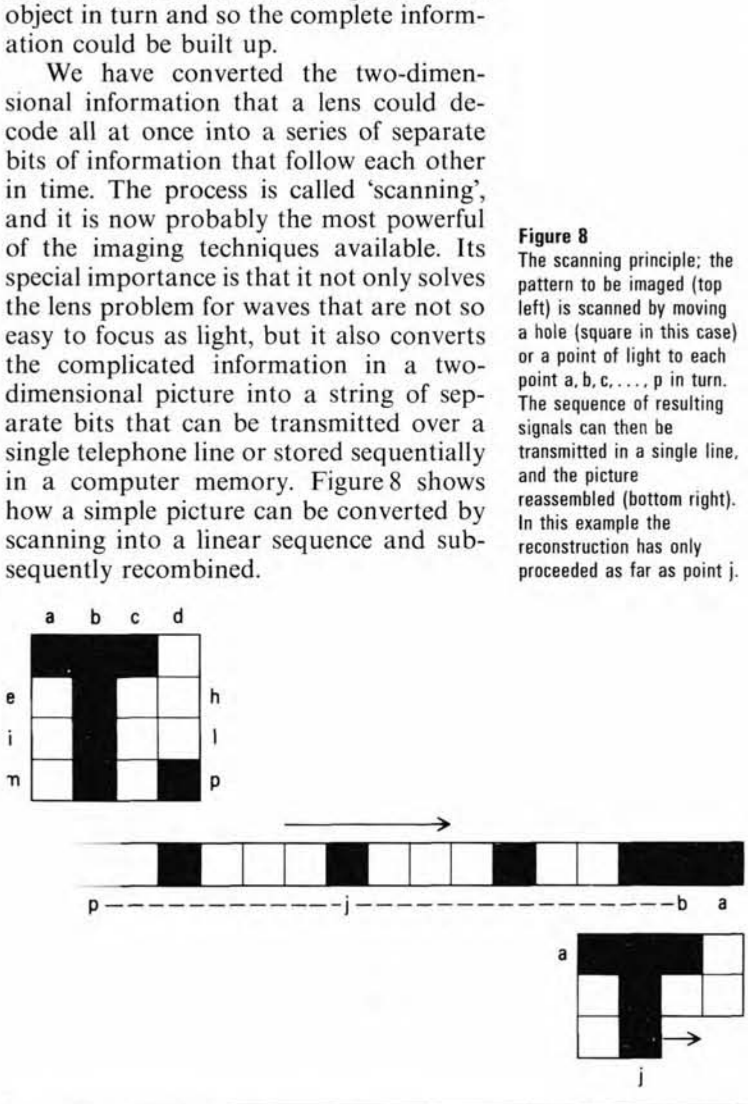



**Figure 8**
The scanning principle: the pattern to be imaged (top left) is scanned by moving a hole (square in this case) or a point of light to each point a, b, c, . . . , p in turn. The sequence of resulting signals can then be transmitted in a single line, and the picture reassembled (bottom right). In this example the reconstruction has only proceeded as far as point j.

An ingenious method of producing images using neither lenses nor scanning is known as 'holography'. We have already met the idea of a hologram; but if the pattern is produced with coherent light and combined with a uniform coherent beam, a special kind of hologram can be recorded on a photographic plate. When the developed plate is viewed in coherent light at the appropriate angle, a full three-dimensional image of the original object can be seen as a result of the recombination of the waves scattered by the holographic record.

But suppose we are dealing with radiation for which lenses, scanning, or the production of coherence to permit holography are all out of the question? The only remaining possibility is to record the pattern of the scattered radiation (that is, the hologram, or fuzzy patch) and to try to interpret it directly. A glance at figure 4(b) will remind you just how difficult this might be. In fact, unless some special conditions are fulfilled it is impossible.

Why should we wish to try? Chemists, metallurgists, molecular biologists, and many other people need to know precisely how atoms are arranged in solid matter. How does graphite differ from diamond though they are both pure carbon? What happens to the atomic arrangement in a metal wire to make it become harder when stretched? Why do some man-made fibres take up dyes more easily than others? However, since the average separation of the atoms is only about $1.5 \times 10^{-10}$ metre, if you look at table 1 you will see that X-rays are the only suitable waves. (In fact, neutrons or protons are sometimes used and, although there are various complications, the basic principles of the interpretation of the patterns is much the same.)

So we have an important group of problems that can only be investigated with waves that cannot be recombined in any of the ways discussed so far. We have already said that direct interpretation of the patch is very difficult, however there are two redeeming features. First, although the X-ray beams cannot, so far, be made coherent enough to create a coherent beam and produce holographic reconstruction, they can be made sufficiently spatially and temporally coherent to make the patch of radiation scattered from an object more like those in figures 4(a) and (b) than the fuzzy patch of the lensless projector. Secondly, there is nearly always some regularity about the arrangement of the atoms in solid matter (that is, most solid matter can be produced in a more or less crystalline form), and so the patterns are more like figure 4(a) than figure 4(b). The patterns are usually called X-ray diffraction patterns and a typical one is shown in figure 9. The experimental arrangements for recording the patterns are fairly complicated because the wavelengths that are conveniently used are not very much smaller than the detail being sought and so patterns must be recorded over a very wide angular range. (Figure 1 will help you to understand why this is so.) However, the interpretation is even more complex.

Earlier in this section I pointed out that we cannot focus an image unless we have some idea of the nature of the object, and it is this fact that enables crystallographers to interpret the diffraction patterns. We do know roughly what size and shape atoms are, we know roughly how far apart they are likely to be, and we usually have some idea of the kinds of arrangements that are likely to be present. Sometimes more or less mechanical ideas of how molecules of a particular shape are likely to pack together can be used, and sometimes we know the structure of a very similar substance and can use that as a guide.

One powerful technique that played a significant part in unravelling the mystery of protein structures is known as the 'heavy atom' technique. It resembles the trick, mentioned in Section 5, of focusing on specks of dust or hairs when trying to get a sharp focus with a slide projector. A 'heavy' atom (that is, with a high atomic number since it is the electron cloud that scatters the X-rays), such as platinum or mercury, is attached chemically to each molecule of the unknown substance; calculations can then be done to 'focus' an image of the heavy atom that is known to be there and may lead to a picture of the whole molecule.

Recently computer techniques have been developed which, in effect, put together all the available knowledge about the structure, and about the size and shape of the atoms, and in a reasonable proportion of problems can produce an image virtually automatically. The same principles of scattering and recombination still apply, though the recombination is theoretical rather than actual, and the limits on precision of the final image also still apply.

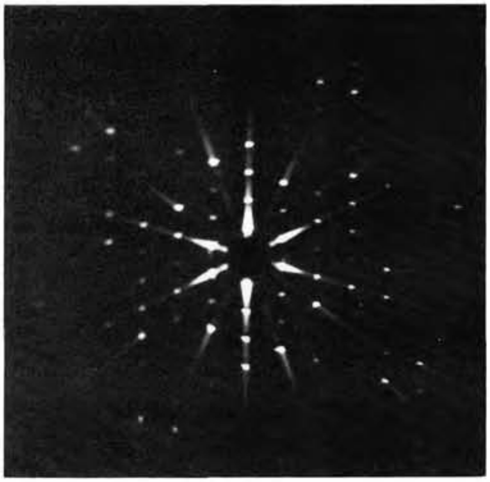


**Figure 9**
Typical X-ray diffraction pattern; this is a so-called precession photograph for the organic substance hexamethylbenzene.
*Professor C. A. Taylor.*

## 6 The nature of the patterns of the scattered waves

We have just been talking about direct interpretation of scattering or diffraction patterns so it is appropriate to say a bit more about the patterns themselves. Apart from changes in scale, and in recording or detection devices, the same principles apply to all radiations, but, because observation is easier, we shall discuss them in terms of visible light.

It is usual to describe the interaction between waves and objects as 'diffraction' and the subsequent interaction between the waves themselves as 'interference' or 'superposition', though text books are far from unanimous about this and specialists in different topics tend to use different words. Strictly, I suppose that X-ray diffraction patterns should be called interference or superposition patterns, and many other logical changes could be made, but once a practice has become established it is difficult to change.

Physicists tend to divide the patterns

produced by the scattering of coherent light into two categories. If the incident wavefronts are planar (that is, the light is parallel) we call it 'Fraunhöfer diffraction', whereas if the wavefronts are any other shape (typically cylindrical or spherical) we call it 'Fresnel diffraction'; we shall discuss only the former. Since the laser beam has fairly planar wavefronts the patterns of figures 4(a) and (b) could be described as Fraunhöfer patterns, but for precise study we use an instrument called a 'diffractometer'. A typical arrangement is shown in figure 10. An object consisting of two holes punched in opaque paper and placed at D will produce a pattern resembling Young's fringes – see figure 11(c). The spacing of the fringes becomes wider as the spacing of the holes becomes smaller and the fringe maxima run in a direction perpendicular to the line joining the holes.

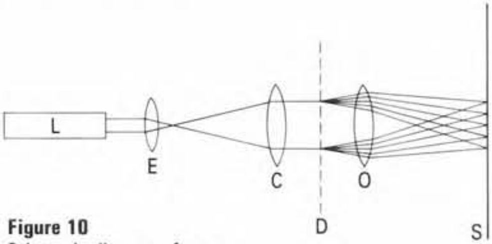
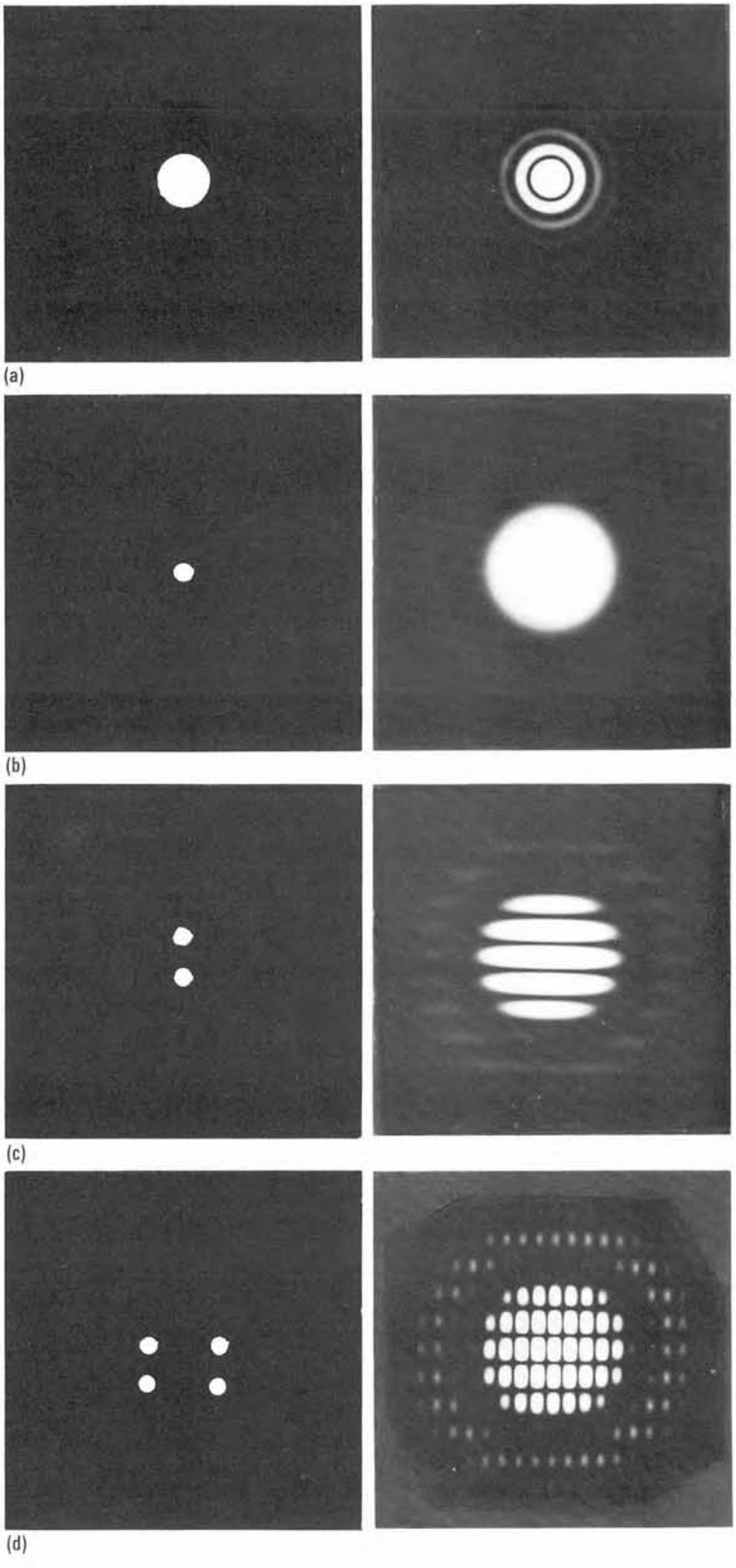
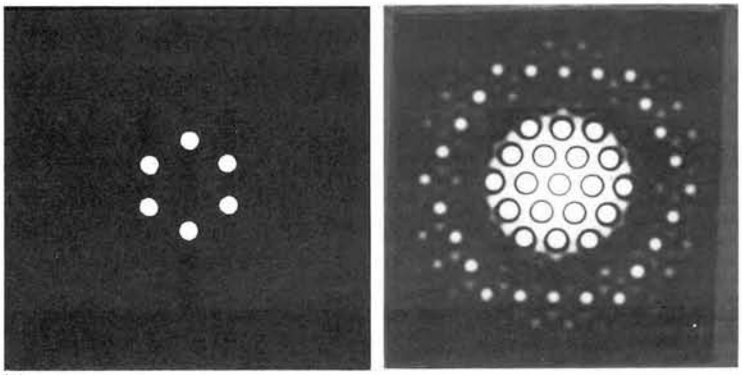
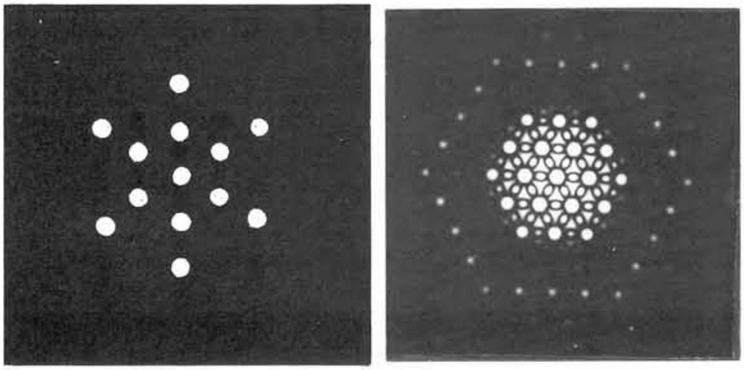



**Figure 10**
Schematic diagram of an optical diffractometer. L is the laser; E is the beam expander (often a microscope objective) to change the parallel beam into a beam diverging from a point; C is a collimator lens to make the light parallel again, but the beam is now of much greater diameter than originally; D is the diffracting object; and O is the objective lens. In the absence of an object at D an image of the effective point source is produced on the screen; and when an object is at D its Fraunhöfer pattern appears on the screen. (Strictly it is not an image of the point source that appears in the absence of an object at D but the Fraunhöfer pattern of the aperture of lens O or C, whichever is the smaller; that is an Airy disc (see discussion in Section 8).)

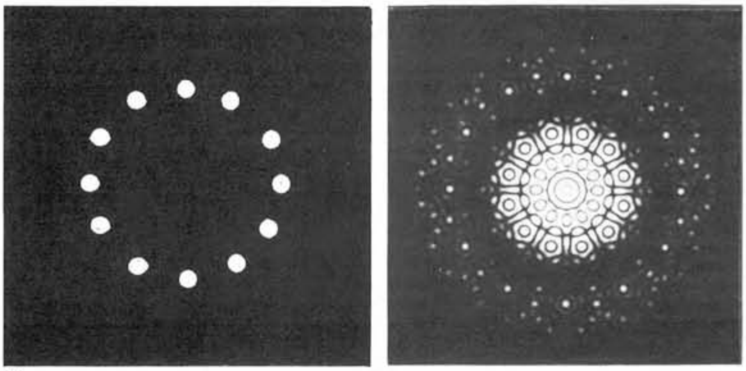




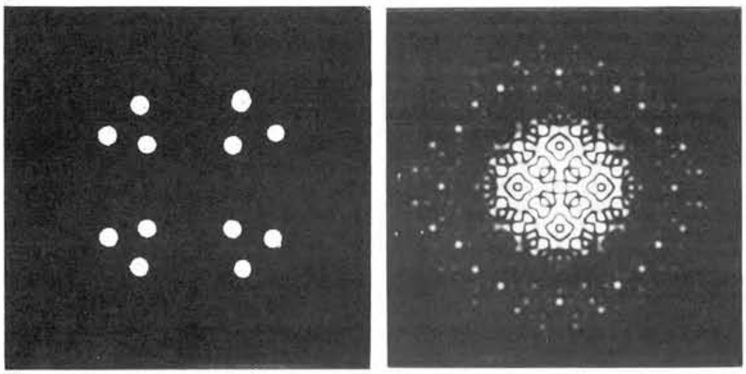
(e)



(f)



(g)



(h)

**Figure 11**
Typical Fraunhöfer patterns produced with a diffractometer like that of figure 10. The lefthand photographs of each pair are effectively the diffracting objects which were a series of opaque screens with one, or more, circular holes in them, and the photographs on the right are enlargements of the resulting Fraunhöfer patterns. (In this case the actual holes used were 1 mm in diameter except for the first which was 5 mm.) *Professor C. A. Taylor.*

Any object can be thought of as made up of a number of pairs of points, and so the pattern of any object can be thought of as the superposition of many sets of fringes of different spacings and orientations. Figure 11 shows how this works out for objects that actually consist of holes, but the principle can be extended to any object if the number of pairs of points is large enough. Notice particularly the pattern for the single large hole, known as an Airy's disc, which plays an important part in the next section.

## 7 What limits the clarity and precision of an image?

We saw in Section 3 that no information about details smaller than the wavelength of the radiation used can be encoded and that, no matter how powerful the rest of the system is, there is just no information to be extracted. This limit is sometimes called the theoretical limit of resolution. There is, of course, no sharp transition as the detail being studied gets smaller, and so the precise definition is a matter of convention. But if we refer back to figure 1 we can see that this theoretical limit could not be reached if the aperture were to be restricted so that waves from a small range of angles only are admitted to the image forming system. Figure 12 is a repeat of figure 1(a) in which the object points are far apart but the aperture is restricted so that there is virtually no phase change from one side of the aperture to the other. This gives rise to an *experimental* limit of resolution that is related both to the wavelength and the dimensions of the imaging apparatus.

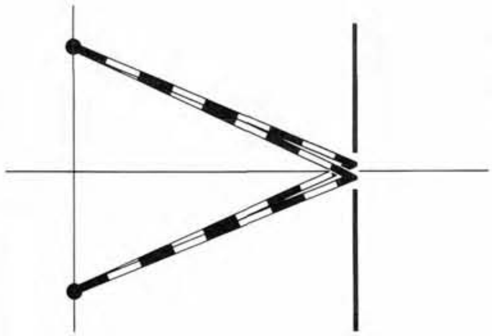


**Figure 12**
Repeat of figure 1(a) showing how the range of phase differences is restricted when an aperture is introduced.

If we remove the screen in the diffractometer illustrated in figure 10 and allow the light to fall on to an auxiliary lens O′, an image of the object at D can be formed on another screen S′ placed to the right of the auxiliary lens (see figure 13). A variable aperture A placed over this auxiliary lens can be used to restrict the scattered light that goes forward to form the image to a smaller range of angles, and thus to imitate what happens in an image-forming instrument when its aperture is restricted. Figure 14 shows the results of such an experiment. Notice particularly that in figure 14(f) there appear to be nine fuzzy holes in the centre block, whereas in the real object – figure 14(a) – we can see that there are really twenty-five! This is another cause of the controversies about details of microscopic objects that were mentioned in Section 5, this time arising from the use of an objective with too small an aperture.

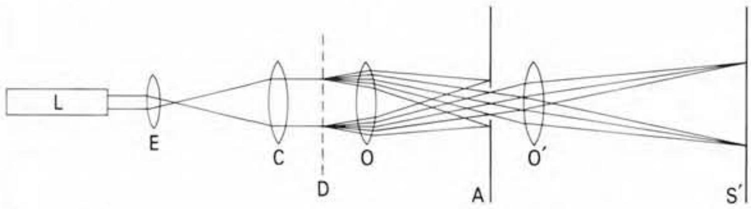



**Figure 13**
In the modified diffractometer the screen is moved out to S′ and an auxiliary lens O′ forms an image on it of the object D. An aperture inserted at A reduces the area of the diffraction pattern that can contribute to the image and this leads to loss of detail and sometimes to false detail.

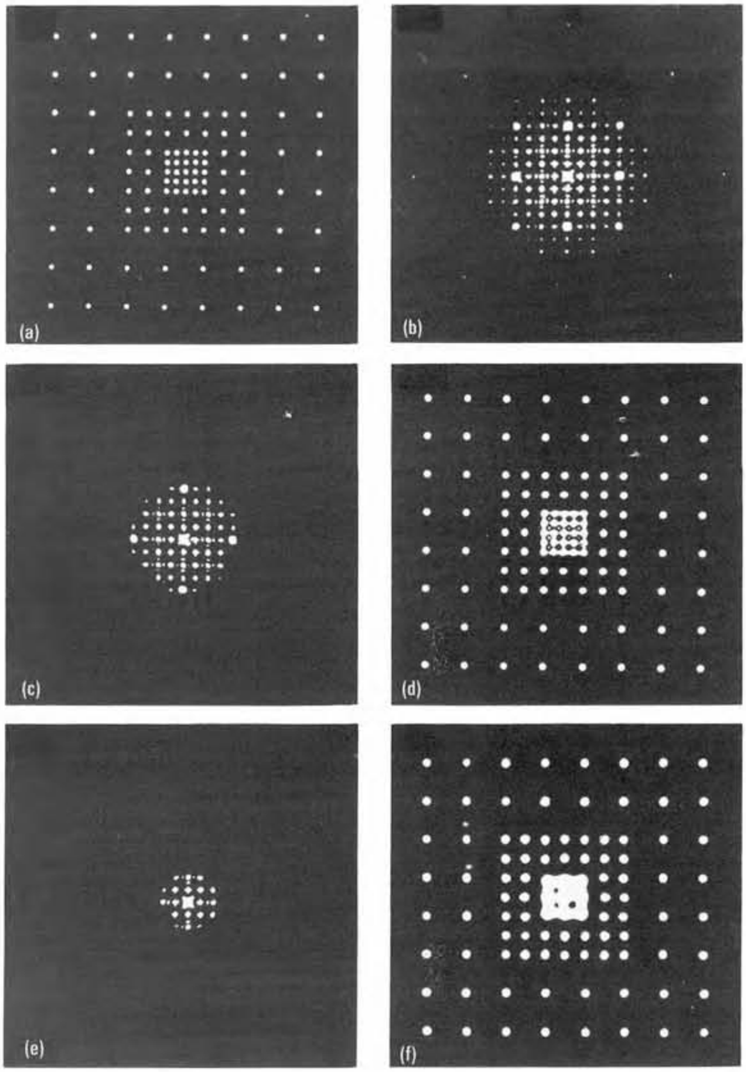
So far we have assumed that the recombination, or focusing, system is perfect, except for its restricted aperture. But if, in addition, there are defects in the lenses (known as aberrations), or imperfections in the scanning or transmission systems, or indeed in any other part of the instrument, then there will be even greater restrictions on the clarity and precision of the image.

## 8 How does this work out in practice?

Optical systems such as microscopes telescopes, and cameras involve the most obvious and direct applications of the principles that we have been discussing. In an astronomical telescope the resolution limit is always the experimental one since, if one is observing points of light from a source millions of miles away it is clear that to accept a full range of angles in order to approach the theoretical limit you would need an aperture millions of miles across! If you think about it, a telescope is an ideal instrument for observing Fraunhöfer diffraction, since a star is so far away that the light entering the system has virtually planar wavefronts. What you see, or photograph, is therefore not an image of the star but rather the diffraction pattern of the aperture of the main lens or mirror.

This pattern will be an Airy disc resembling the pattern of figure 11(a). The smaller the telescope aperture the larger will be the ring pattern and so the less likely you are to be able to distinguish stars that are close together. This is an alternative way of approaching the idea of the experimental limit of resolution. Of course if you are looking at a large object such as the Moon

**Figure 14**
The effects on the reconstructed image of restricting the aperture. The modified diffractometer of figure 13 was used.
(a) is the diffracting object placed at D and (b) the Fraunhöfer pattern at A.
(c) is a repeat of (b) but with a circular aperture placed over it in the plane of A.
(d) shows the reconstructed image produced when the auxiliary converging lens O′ is introduced to collect the light passing through the aperture and recombine it to form an image at S′.
(e) and (f) show the results when an even smaller aperture is used.
*Professor C. A. Taylor*




you *do* see an image; but each point of the image that ought to be a mathematical point becomes an Airy disc, and so any fine detail in the image will be blurred to a greater or lesser extent depending on the aperture in use. The same principles apply to radio telescopes.

It is possible to design an objective lens for an optical microscope that will take in almost the full range of angles up to 90° on either side of the axis, and so very closely approach the theoretical resolution limit.

The effect of aperture restriction is particularly important in the case of electron microscopes, where the effective wavelength is far smaller than would be needed to image atoms successfully. The fact that the electromagnetic lenses used to focus the images are very difficult to make with apertures of a reasonable size the experimental limit, both from the diffraction standpoint and also because of aberrations and other imperfections, has meant that until recently it has not been possible to image individual atoms. Even now such images are possible only with special electron microscopes and certain types of materials.

There are two main types of electron microscope. The first follows closely the design of an optical microscope but with electromagnetic lenses; the second type uses a fine beam of electrons to scan the surface of the object being studied, and the image is built up point-by-point on the screen of a cathode ray tube. On the face of it we might expect the scanning microscope to escape the resolution limit but this is not so; the resolution is determined by the size and shape of the scanning spot and the aperture and aberration problems apply just as much to producing the fine spot as they do to producing a sharp image in the more conventional way.

In radar systems the aerial array, or paraboloidal dish, focuses a narrow beam which is scanned over the object and information about the distance of the object is provided by the time delay between the pulse of radio waves being sent out and being picked up by the receiver. This time delay measurement is obviously a measurement of phase difference and, as we have seen, you would not expect to be able to measure phase changes produced by objects much closer to each other than about a wavelength. The precision of the information about the direction of a particular object from the radar equipment is more likely to be limited by the aperture of the transmitter, defects in the system, and so on. These all set a limit to the narrowness of the scanning beam that can be produced, which in turn determines the limit of directional resolution. As with most imaging systems, the interpretation of the resulting picture can need a great deal of skill, and 'focusing' again depends on knowing the kind of object being viewed.

Ultrasonic systems for safely imaging the inside of the human body (for example, in the examination of pregnant women) depend on virtually the same principles as radar. Very high frequency sound waves focused into a narrow beam scan the object and reflections from the various kinds of tissues and fluids present send back phase information that can be used to build up a picture on the screen of a cathode ray tube. Again the limitation on resolution is experimental, as the frequency used (up to $10^8$ Hz) gives wavelengths down to $3 \times 10^{-6}$ m, which is far smaller than the kind of detail to be imaged.

In recent times digital imaging has increased in significance. Once an image has been turned into a linear sequence of bits of information by scanning, it is easy to convert these into digital form and hence store, transmit, or process the information in some way before reconstructing the image as in a normal scanning process. Why should this be an advantage? There are two main reasons. First, all signalling systems, whether aural or visual, suffer from the effects of noise, that is from unwanted information that confuses the signals. We are all familiar with the hiss that can swamp a radio signal when we are near the edge of the reception area, or with the snowstorm effect on a television screen when the signal is not tuned properly, or the aerial is pointing in the wrong direction. Both are examples of noise. But if the signal is transmitted in the form of binary digits where all that one needs to know is whether there is a signal or not, and the intensity of each signal is immaterial, then the effect of the noise can be minimized. Secondly, and perhaps more importantly, once a signal is in digital form all kinds of computer processing are possible. The very faint contrast in an X-ray picture can be enhanced to give the surgeon a clearer picture; pictures from satellites or space ships can be 'cleaned up' by using the principle that the picture you are seeking is likely to be constant, whereas all the interfering signals are likely to be random; and, of course, on a more trivial, but very lucrative, commercial level, all the extraordinary modifications of texture, shape, and colour used in video tapes and on television depend on the same principles.

But throughout all these various applications the five basic principles that we have spelt out always apply whenever radiation is used to form an image. You may not immediately recognize their impact, but it is a very good exercise to see for yourself how they apply to any system with which you are familiar.

1 The first stage in an imaging process is the scattering of the waves by the object (or radiation of the waves if the object is self-luminous).
2 The second stage is recombination, or focusing, of the scattered waves. (This can also be described as decoding of the encoded information.)
3 Focusing or decoding is only possible either if every single parameter of the system is known accurately, or, as is more usual, if you have a fairly good idea of what the object looks like.
4 Details smaller than about a wavelength cannot be encoded and hence cannot be imaged.
5 Even if the wavelength is suitable, information can only be extracted reliably if the experimental system is free from defects and is of sufficiently large aperture to take in the full range of waves scattered by the object.

# Our nuclear history

**GEORGE MARX**
**Department of Atomic Physics, Eötvös University, Budapest**

Professor Marx is a Hungarian nuclear physicist who has done research on fundamental particles and high energy physics. He has also taken a lively interest in teaching physics at school and university. In this article he shows how the difficult decisions about energy which we face at the end of the twentieth century can be seen as a consequence of the history and evolution of the Universe and of life and Mankind's activities on our planet. Ideas from many parts of the Nuffield A-level physics course are used here, and you may want to read the article a couple of times: once when thinking about fuel supplies and again after you have studied ideas about entropy and wave mechanics at the end of the course.

CONSERVE ENERGY! From time to time public campaigns encourage us to 'Save it!', 'Turn it off!', 'Turn down the thermostat!'; we are told that 'Gas is too precious to waste – use it wisely'. And yet, according to physics, energy *is* conserved! So what's the fuss about?

Power stations are often built on sea shores and at river sides; elsewhere we see giant cooling towers, which are erected just to dissipate the heat produced. Why is the energy – obtained from expensive fuel – wasted deliberately? This is a very good question. If you are interested in understanding the role of energy in our society today, a good idea is to take a historical look at the problem.

Our bodies and all the everyday objects on Earth, the planets and the stars, are made of atoms which in turn are composed of smaller particles. The Universe is a closed system of a tremendous number of particles. It has an inexhaustible number of degrees of freedom; it can redistribute energy among its atoms, but it cannot obtain energy from 'outside'.* It has to live and work with the existing amount of energy. What happens both in free nature and in human affairs is that sooner or later any concentrated energy gets dispersed among the degrees of freedom available in the environment. Energy is conserved (First Law of Thermodynamics), but it is dissipated spontaneously (Second Law). If Man is interested in creating order, in concentrating energy (to produce locomotion, to perform work, to process information in our daily life), he has to pay a price for this unusual course of events: somewhere else more energy must be scattered over the degrees of freedom of the environment. This is the only way human society can reach its technical goals because it cannot violate the laws of nature. And the history of the Universe can also be understood by this spontaneous random distribution of energy among all the particles. It is worth observing faraway stars, to learn about the immense cosmic past, because the knowledge obtained will help us understand the present energy crisis and the prospects for the future.

* A *closed system* is one which neither gains nor loses energy (or matter) from elsewhere. The number of *degrees of freedom* a system has is a measure of the number of ways in which energy and matter can be distributed in the system.

## The origin of the Universe: the Big Bang

The Second Law, giving a direction to time, was discovered in the last century, when engineers became concerned with constructing better and better steam engines. But formulating the Second Law produced another problem. In a closed universe total disorder should be the natural result of events. (Clausius called this disappearance of thermal, density, and chemical differences 'heat death'.) How is it, then, that we still experience heat and cold, light and darkness, decay and birth?

The answer to this question has been understood only in our century. The answer begins with a curious truth about dynamics.

The Moon cannot stand still in the sky. A stone cannot float in the air, it must fall down to the Earth, or – if it possesses enough kinetic energy – fly away. The laws of motion do not have an ultimate static solution! This theoretical conclusion of Alexei Friedman (USSR, in the early twenties) was confirmed by Hawking and Penrose (UK, in the sixties) in full mathematical rigour. In the 1920s Hubble (USA) observed that the galaxies are flying apart: the Universe is expanding. By extrapolating back into the past, one reaches an infinitely high density about 15 billion ($15 \times 10^9$) years ago.

Because the particles were compressed tremendously at that time, one would expect them to have formed the most stable nuclei (metals like iron). On the other hand, the present Universe is actually dominated by the lightest elements – hydrogen, helium. The only explanation can be (as Gamow argued in the forties) that the early Universe was hot. The wild thermal motion prevented the formation of composite nuclei in the early dense era. But later, in the expanding clouds of cosmic matter, as mutual gravitational attraction decreased, so the increase in gravitational energy consumed kinetic energy of expansion and thermal motion*. Matter was cooled by expansion, just as the rising hot air above warm land cools down to produce hail in high summer†.

The leftover glow from the early hot Universe was discovered by radio engineers of the American Bell Laboratories in the sixties, when they were scanning the sky for quiet radio bands for satellite telecommunication. They observed an overall radio noise, corresponding to a radiation temperature of $-270°C = 3$ K. From this present temperature it has become possible to reconstruct the course of events*.

The shape of the spectrum of this radiation corresponds to that of the Planck radiation law. This indicates that the early hot Universe was in a state of complete disorder. The 'heat death' of Clausius happened, but in the past! In the very first second of its existence, the Universe had a temperature above $10^{10}$ degrees Celsius. At such a temperature no composite nuclei existed, and free protons and free neutrons were present in equal numbers. They transformed into each other in violent collisions.

## The formation of the light elements

After the first second of the life of the Universe thermal motion lost its violence, the average kinetic energy dropped, and collisions were not energetic enough to produce more neutrons. The existing neutrons (being a bit heavier than protons) decayed gradually†. But before they all had time to decay, some of the protons captured neutrons: the first composite nuclei, like deuterium (a heavy isotope of hydrogen), helium, and lithium were formed by fusion, all in the first three minutes.

But spontaneous radioactive decay of free neutrons terminated this fusion chain before it reached an equilibrium composition. Since nuclei have a positive charge their mutual electric repulsion prevented any further fusion in the cooling gas: nuclear matter had been trapped in the Coulomb potential wells of light nuclei. From the primordial particles only the first few elements of the Periodic Table were produced. The expansion and the conse-

* The gravitational potential energy of a collection of particles is negative. As the distance between particles increases the gravitational potential energy becomes less negative, that is it increases. This increase is paid for by a decrease in the kinetic energy of the particles.

† When hot air rises it expands and the work done in expansion causes a cooling. Other examples are the sudden expansion of the compressed gas from a $CO_2$ cartridge or a bicycle tyre.

*The wavelength of radiation emitted by a body depends on its temperature. The Sun, with a surface temperature of 6000 K, radiates mostly visible and infra-red with a peak at about 475 nm. A body at 3 K radiates mostly in the mm–cm wavelength region.

†The mass of a neutron is 1.001 times the mass of a proton. *Free* neutrons are unstable and decay to form a proton plus an electron (plus a tiny particle, the anti-neutrino: $n \longrightarrow p + e + \bar{\nu}$). The decay is slow – the half-life is about 1000 s. Neutrons in a nucleus are stable.

quent cooling was so fast that nuclear matter was not able to keep up with the change of conditions: it retained its 'hot' composition in spite of the 'cool' environment (like supercooled water which keeps its liquid state below the freezing-point).

Millions of years passed by before the temperature dropped below a thousand degrees. Thermal motion calmed down, protons captured electrons and formed hydrogen atoms. The Universe was filled by luke-warm hydrogen gas (a good first approximation for cosmic matter even today), contaminated by helium and by tiny traces of other light atoms (deuterium and lithium did not amount to more than $10^{-5}$ per cent). The gas, now cool, became unstable against local gravitational contractions. It condensed into dense hydrogen clouds, the clouds fragmented into galaxies, then the galaxies to stars. The work of gravity heated the contracting gas so that it glowed and shone. The darkness of space was illuminated again, now by starlight.

The work of gravity increased the central temperature in stars and temperatures above a million degrees again existed. Proton collisions became more and more violent. Occasionally, quantum tunnelling helped a fast proton through the repulsive potential wall of another.* The proton–proton pair dissociates very rapidly; occasionally the transient conglomerate suffered a radioactive decay before coming apart:

$$p + p \longrightarrow (pp) \begin{cases} \nearrow p + p \\ \searrow D + e^{+} \end{cases}$$

By emitting a positive electron, a deuteron (D) was formed. This bound state of a proton and a neutron is the nucleus of deuterium, heavy hydrogen. Its nuclear binding energy was radiated†, feeding and prolonging the shining of the star. But deuterium nuclei are not long-lived at temperatures of several million degrees: they merge into helium nuclei.

$$D + D \longrightarrow He$$

(in detail: $^{2}H + {}^{2}H \longrightarrow {}^{3}He + {}^{1}H$;

$^{3}He + {}^{3}He \longrightarrow {}^{4}He + {}^{1}H + {}^{1}H$)

The helium nucleus possesses a closed shell (see page 44), it is stable even at several million degrees*. In most stars of our night sky this hydrogen–helium build-up feeds the central power station, supplying stellar radiation for billions of years. The steady flow of nuclear matter towards its equilibrium state is retarded by the Coulomb barrier surrounding the positively charged nuclei. Even now three-quarters of the Universe is made of primordial hydrogen, one-quarter of helium (a second approximation for the present composition). Our Universe is still young from a nuclear point of view.

## Carbon and oxygen are formed

When the hydrogen supply of the central core of a star gets exhausted, thermal fusion slows down. The star shrinks and the liberated gravitational energy makes up the difference between the energy from fusion and the energy radiated. The star heats up to 100 million degrees. The larger impact speeds needed to cross the higher barriers separating the nuclei of greater charge are reached. At this high temperature even triple collisions become frequent. So the nuclear build-up goes further:

$$3He \longrightarrow C \qquad He + C \longrightarrow O$$

Oxygen-16 has a closed shell structure, at which the fusion chain terminates again.

* Quantum tunnelling is a process in which a particle may penetrate a potential barrier even though by classical physics it doesn't have enough energy to do so. In this case two protons approach each other closely enough to form a transient (pp) in spite of the repulsive electric force between them.

† The energy of the two bound particles is less (more negative) than their energy when far apart. The difference in energy, equal in magnitude to the nuclear binding energy, is released as radiation.

* Nuclei having 2, 8, 20, 28 ... protons (or neutrons) are more stable than others. This can be explained in terms of a series of nuclear shells each of which can accommodate a fixed number of particles. A closed shell means a very stable nuclear structure. The idea is similar to the idea of the electron shells which explain the stability of certain atoms, and both electron and nuclear shells have their explanation in wave mechanics. But the details are different, and the numbers of particles in closed nuclear shells are not necessarily the same as those in the closed electron shells of the inert gases.

Red giant stars (like Antares and Arcturus glowing red in the summer sky or Betelgeuse in the winter sky of the Northern hemisphere) produce carbon and oxygen from helium. The stellar wind, blowing outwards from the star's surface contaminates space with these life-essential elements to one part per thousand (a third approximation for the present composition of the Universe). Heavy elements like the metals are not present in any practical quantity because the stars do not yet have time to develop temperatures high enough to make the penetration of the higher potential barriers around oxygen nuclei possible.

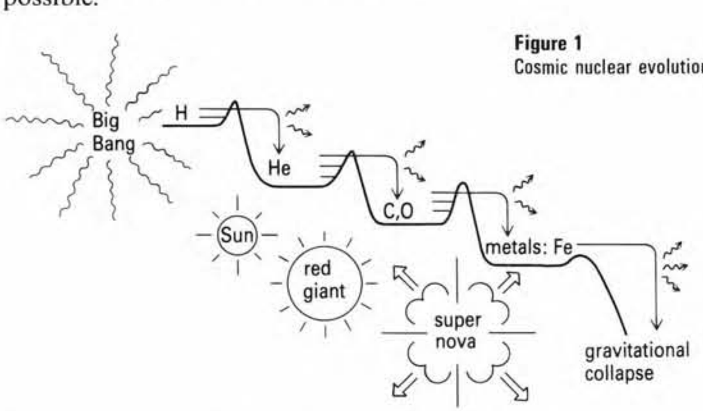



**Figure 1**
Cosmic nuclear evolution.

## Supernovae and nuclear equilibrium

Sun-sized stars stay in a hydrogen-burning state longer than the present age of our Galaxy. But any considerable overweight results in faster ageing: a star of ten solar masses radiates more lavishly, and exhausts its hydrogen and helium supply within a billion years. In the critical period of diminishing nuclear fuel the star shrinks at an uncontrolled speed, the work performed by gravitational pull heating it up to billions of degrees when all possible nuclear reactions can occur, and so an equilibrium population is produced along the Periodic Table of chemical elements. The most abundant isotopes are iron and the neighbouring metals. At such a high temperature not only the states of minimum energy are populated, but, in smaller amounts, states with high energy content are also present: light nuclei and very heavy ones as well, like uranium and thorium (as follows from the Boltzmann distribution*). When the end state of nuclear evolution, the equilibrium state, is reached, no more reactions occur. The death of massive stars arrives with a collapse. The liberated energy gives rise to a shock wave, which strips the outer stellar layers off. The rapidly expanding stellar envelope raises the brilliance of the dying star for a few weeks: astronomers register this giant flash as a supernova explosion. A famous one was the Crab Nebula, visible by day for some weeks in 1054. The ejected metal-rich gas cools down and surges through the neighbouring interstellar gas and dust. New density concentrations are formed, bearing a second generation of stars. These new stars are not only contaminated by carbon and oxygen, but also contain 0.01 per cent metals (a fourth approximation of the present cosmic abundance of chemical elements). Our Sun was born in this way less than five billion years ago.

## Formation of planets

If the star-forming cloud was whirling, its angular momentum prevented its shrinking to a single star. The outer fragments which formed planets took over most of the angular momentum by their orbiting and spinning motion. Only the birth of planets enabled the central mass to become a star.

The most abundant elements in the planetary material were hydrogen, helium, oxygen, and carbon, with traces of metals and other elements. Consequently, the most common compounds in the gas and dust were, in decreasing order of abundance, $H_2$, the inert gases, $H_2O$, $CH_4$, $CO_2$, and metal oxides. This composition can be found in the cold outer planets of the Solar System. In the luke-warm inner region, however, the hydrogen and inert

* In a collection of particles some will have high energy, some low. The number of particles having a particular amount of energy is given by the Boltzmann distribution (see Unit K, 'Energy and entropy' of the Revised Nuffield Advanced Physics course). As the temperature increases a larger fraction of the particles have higher energy, which is why, for example, hot water evaporates more quickly than cold and why the conductivity of semiconductors increases with temperature.

gases never condensed, but were blown away by solar radiation and the solar wind. The metal compounds built planetary cores, surrounded by atmospheres of water, methane, and carbon dioxide.

The energy released by fast decaying radioactive elements in the hot furnace of the supernovae melted the body of the planet. This melting produced a separation by density. Within a billion years most radioactive isotopes decayed, leaving only the long-lived uranium and thorium. Slowly a rigid crust was formed. We experience the delayed heat of the ancient supernovae – the release of energy from radioactive nuclei formed at that time – in volcanoes and hot springs. Oceans of water precipitated (thanks to hydrogen bonds formed between the polar $H_2O$ molecules). Oxidized planets ($H_2O$, $Fe_2O_3$, etc.) with carbon-rich atmospheres of $CH_4$ and $CO_2$ were left. This chemical equilibrium can be observed on Venus and Mars. But the fate of our mother planet turned out to be very different.

## Development of life on Earth

Earth does not behave like a closed system – it gains energy from the Sun. The star called the Sun is made mostly of primordial hydrogen. This solar material leaked slowly through potential tunnels towards deeper 'energy valleys', the energy set free providing sunshine. The energy of the ultra-violet rays of the Sun was enough to break bonds in the hydrogen-rich molecules on the young Earth and in interplanetary space. As bonds were broken, new compounds were formed in chemical reactions:

$$\text{light} + H_2O \longrightarrow H + OH$$
$$\text{light} + CH_4 \longrightarrow CH_3 + H$$
$$CH_3 + OH \longrightarrow CH_3OH$$
$$\text{light} + CH_3OH \longrightarrow CH_2OH + H$$
$$CH_2OH + OH \longrightarrow H_2C(OH)_2 \longrightarrow H_2CO + H_2O$$

$H_2CO$ (methanal, formaldehyde) is a polar molecule, soluble in water. Rains carried it into the protective ocean (figure 2).

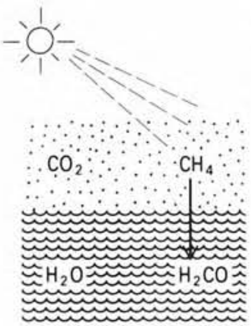



**Figure 2**
The young Earth.

The methanal molecule contains a double bond ($H_2C{=}O$) which makes it reactive, so methanal was polymerized to glucose:

$$6H_2CO \longrightarrow C_6H_{12}O_6$$

The high energy bonds of sugar conserved a fraction of the absorbed solar energy*. In this way sea water had become a nourishing soup, raised out of its chemical equilibrium by the rays of the Sun. In this soup living creatures were formed in a surprisingly short time (within a quarter of a billion years). They fed on this broth and spread fast. Soon they ate up all the nourishing compounds in the ocean. In the following lean years some starving organisms invented a way to harvest light quanta of lower energy (longer wavelength) which penetrated the sea water. This was used to break bonds in $CO_2$ and $H_2O$; hydrocarbons (oil) and carbohydrates (sugar, starch, cellulose) and later coal were formed (figure 3).

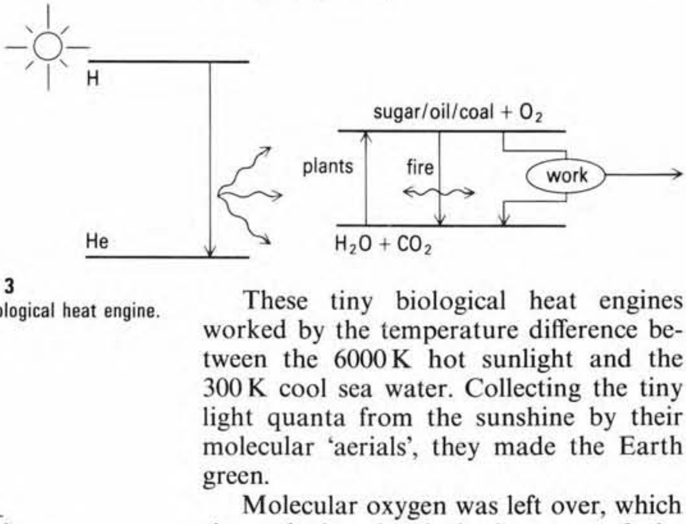



**Figure 3**
The biological heat engine.

These tiny biological heat engines worked by the temperature difference between the 6000 K hot sunlight and the 300 K cool sea water. Collecting the tiny light quanta from the sunshine by their molecular 'aerials', they made the Earth green.

Molecular oxygen was left over, which changed the chemical character of the atmosphere.

* The term 'high energy bond', though a popular one, is perhaps a misnomer. There is no bond in the molecule which – like a coiled spring – is waiting to release energy. Energy is always needed to break a chemical bond. But if a new compound is formed, and if more energy is released in the formation of the bonds in it than was needed to break the bonds in the original compound, then a net amount of energy will have been released in the reaction.

## Mankind and the use of fuels

Combustible organic compounds in the sea, in the grass, in the woods and the oxygen-rich atmosphere above them built up a chemical tension, unprecedented in the Solar System. Some adaptive creatures learned to exploit this chemical tension: by feeding on organic fuel and by burning it at an accelerated rate in the oxygen-enhanced atmosphere they were able to realize a more intensive way of living: locomotion, predatory behaviour, fast evolution. The animal kingdom succeeded in inventing more and more efficient regulatory and respiratory systems. In less than a billion years a socialized animal called Man emerged. By his social system Man explored the natural resources of energy with an unprecedented intensiveness. Man invented fire for cooking and heating, the steam-engine and internal combustion engine for industry and transport.

Before the industrial revolution, Man made use of food and wood to get energy. This meant that a small fraction of the actual power of the thermonuclear reactor inside the Sun was utilized, the solar energy having been transformed into chemical energy by biological photosynthesis.

The industrial revolution increased the public demand for energy to drive engines. Man started mining coal, oil, and gas. Mankind consumes fuels collected by the green biosphere from solar radiation over millions of years. The worldwide industrial revolution is fast enough to empty the terrestrial reservoirs of chemically stored energy within a century. The end of fossil fuel supply has come within sight.

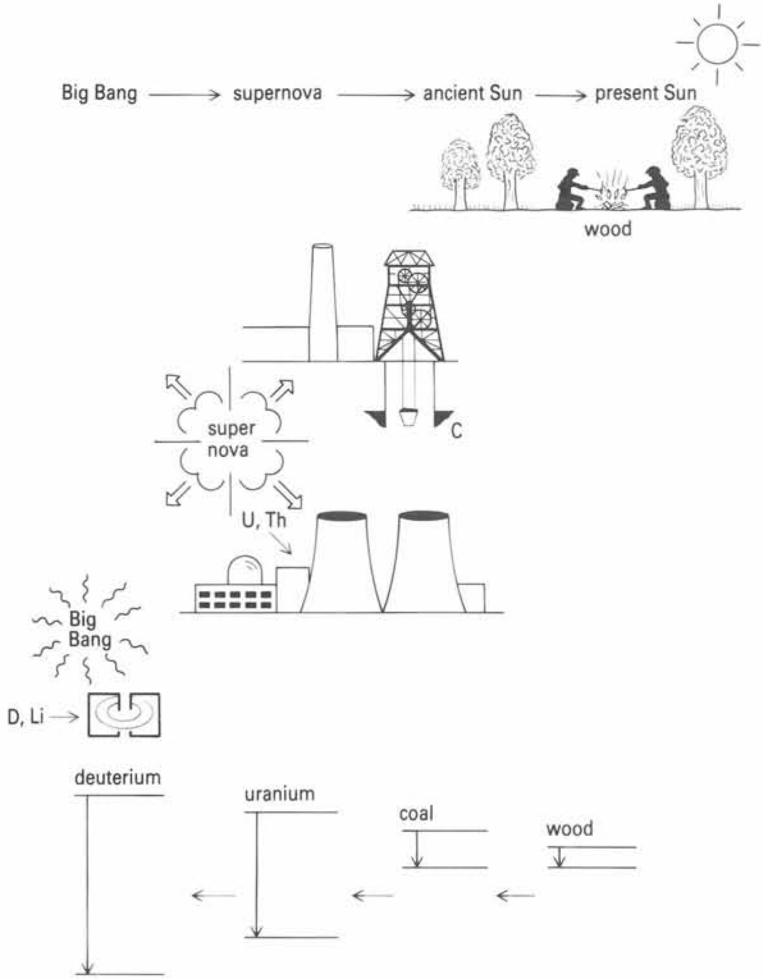



**Figure 4**
Digging deeper and deeper in time for free energy.

Taking our present energy hunger into account, society faces a dilemma. Either we go back to the use of soft energy offered by such renewable sources as wood, wind, waves, rivers (that is by sunshine), accepting the fact that this thinly distributed energy enables only a decentralized 'green' rural way of living for a small population, as in the pre-industrial era. Or we insist on the centralized distribution of energy through a network of wires and pipelines (needed for an urban life-style), and consequently to exploit deeper reservoirs of hard energy (figure 4). Besides the limited supply of coal, oil, and gas, a possible option is offered by the energy of the ancient supernovae, trapped in the nuclei of uranium and thorium. These nuclei are canned sparks with a temperature of several billion degrees, but their seals are – luckily – hard to break. Man has already learned how to make bombs with them, but he has also learned to build fission reactors, and these offer a way to survive the energy-poor decades ahead. Coming generations have to live with nuclear power stations, with radioactive waste, in the shadow of nuclear weapons.

The controlled fusion of light nuclei is a great promise of clean energy for the next century. The light elements (like the heaviest ones) are far from nuclear equilibrium. Deuterium and lithium have preserved the very high temperature of the first three minutes of the Universe due to the slow rate of nuclear evolution compared to biological and social progress. Man learns how to dig deeper and deeper into the past for hotter and hotter sparks; Man invents newer and newer technologies; Man uncovers higher and higher thermodynamical tensions, to get more available energy. Each is harder to get; each is more frightening to have!

Our society has now got out of equilibrium. The fast progress of science and technology on the one hand, the slow development of public understanding and human morals on the other, have resulted in a social tension which produces violent discharges from time to time. This social tension has increased to a dangerous level. We seem to be faced with the choice: zero growth or responsible progress? As H. G. Wells stressed in his *Short History of the World*: 'The future is a race between education and catastrophe'. The old may offer education, but it is the young who will have to avert catastrophe.

## Appendix: the nuclear valley

For many people the nuclear world is an alien, unknown, and consequently fearsome territory. On Earth nuclei are just dormant centres of force, hiding their internal energy for billions of years. But in the Universe nuclei provide the driving force of cosmic history. In this century Man has learned to understand the physics of nuclei. And it turns out that it is simpler to understand how nuclei behave than to explain complex molecules or solids or bacteria.

Nuclei are built from two types of particles: protons and neutrons. The proton has a positive electric charge, the neutron is neutral. Both are about 2000 times heavier than the electron. The neutron is slightly heavier than the proton. This makes the free neutron unstable against radioactive decay into a positive proton and a negative electron.

Protons and neutrons interact with each other through nuclear forces. The fundamental properties of this force have been learned from scattering experiments. Collisions of protons and neutrons have shown that the nuclear force is strong, has a short range, and is not affected by the electric charge of the particles.

Protons and neutrons are held together by this strong force in the nucleus. The simplest bound structure is the nucleus of heavy hydrogen (deuterium) made of a proton and a neutron. Due to the short range of the nuclear force, the size of the deuterium nucleus is necessarily small. From wave mechanics we know that a particle confined to a very small region of space has a large kinetic energy. The kinetic energy is almost as large as the absolute value of the nuclear potential energy (figure 5). So – in spite of the strength of the nuclear interaction – the deuterium nucleus is a loose structure, compared with other nuclei. This makes it a valuable nuclear fuel. (A proton and neutron are

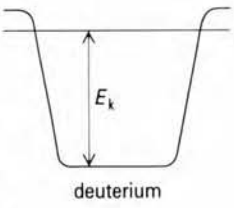



**Figure 5**

just able to form this loose bond, but two protons have no bound state, due to secondary effects, including the electric repulsion between the two positive charges. The non-existence of the proton–proton bound state slows down nuclear build-up in Sun-like stars from microseconds to billions of years.)

Light nuclei can be described as particles enclosed in a tiny box. Wave mechanics teaches us how to relate the energy levels of a nucleus to the standing wave patterns associated with each nuclear particle. The symmetry of standing wave patterns and the rules about the number of particles in each give light nuclei a definite shell structure (figure 6). (Similar rules about electron standing waves give rise to the energy levels of atoms, electron shells, and the Periodic Table.) The first nuclear shell closes at helium: two protons and two neutrons in the lowest s state. (This is why helium is the end of nuclear build-up in Sun-like stars.) The second shell closes at oxygen: the lowest s state and all the three lowest p states are filled by protons and neutrons. (This makes oxygen the end of the nuclear reactions in red giant stars.) On the other hand, lithium is an open structure, like alkali metals in chemistry. (This makes it a reactive nuclear fuel.) See figure 7.

Rutherford-type scattering of charged particles shows that the heavy nuclei have a diameter which increases with the cube root of the number of particles forming the nucleus. This means that the volume of the nucleus is proportional to the number of particles it contains; consequently the density of heavy nuclei is contant along the Periodic Table, just as the density of water droplets is independent of their size. This observation suggests the use of the droplet model for the description of heavy nuclei. If two water droplets (or mercury droplets) merge, a smaller fraction of the constituent particles will be on the surface with unsaturated interaction capabilities, so the potential energy decreases when drops merge. A similar phenomenon occurs in the case of nuclei. The common cause is the constant density and the fact that the range of the interaction is less than the diameter of the droplet. The potential

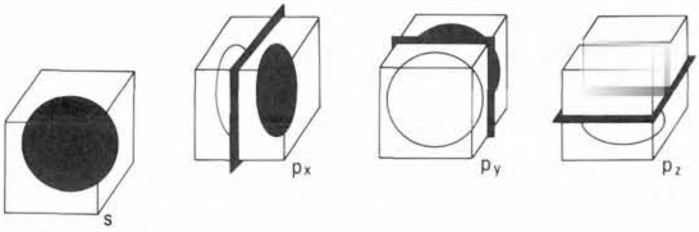



**Figure 6**
Standing waves in a box.

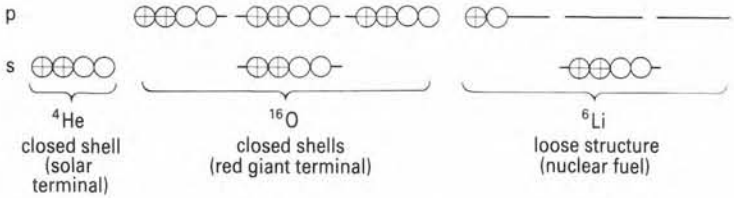



**Figure 7**
Shell structure of light nuclei. (⊕ represents a proton, ○ a neutron.)

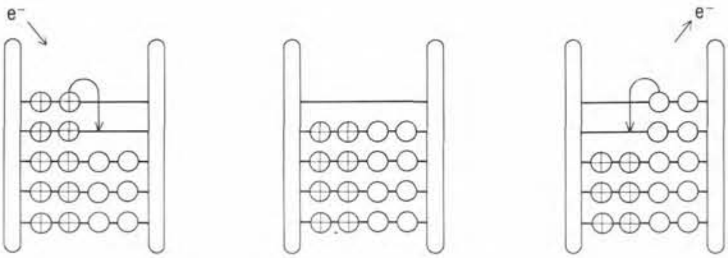



**Figure 8**
The Pauli walls.

energy per particle decreases (becomes more negative) as the number of particles in the nucleus grows. Let us call this deepening of the average binding energy per nuclear particle with increasing mass number the Yukawa slope*. The nucleus contains $Z$ protons and $N$ neutrons. They populate the charge independent energy levels in pairs, so for a given $Z + N$ nucleon number the minimum of the kinetic energy is at $Z = N$. If there are more protons than neutrons, or vice versa, there will be a sharp increase of kinetic energy, leading to the Pauli walls on both sides† (figure 8).

*H. Yukawa (1907–1981), a Japanese physicist who put forward ideas to explain the force between nuclear particles.

†W. Pauli (1900–1958) stated the principle that limits how many particles can occupy the same energy level.

On the other hand, a large number of protons produces a high repulsive Coulomb energy, proportional to $Z^2$. Let us call this $Z$ dependence of the energy per particle the Coulomb slope.*

In this way one can orient oneself in the nuclear valley on the $Z$, $N$ map. This valley looks like a narrow canyon, in which nuclear matter flows from each direction towards the deepest point: towards the iron sea (figures 9, 10). (The potential energy per particle is lowest – most negative – in the iron nucleus.) This flow towards iron drives the nuclear history of the Universe. The slow flow of nuclear matter to deeper and deeper levels is called radioactivity. The decay energy dissipates in the environment outside the nuclei so this is a one-way motion – there is no flow out of the iron sea.

If one has a look at the actual cosmic distribution of nuclear matter in the Universe (figure 11), it becomes evident that our Universe has not yet reached the equilibrium distribution corresponding to the low 3 K of the background radiation. It looks like a canyon a few minutes after a heavy rain storm. Coulomb barriers slow down the flow towards the iron sea, preventing the fast dissipation of energy. Most of the nuclei are still on the Yukawa slope (light elements); their fusion feeds starlight.

The loose deuterium and lithium nuclei may feed the thermonuclear power stations of coming generations. A tiny

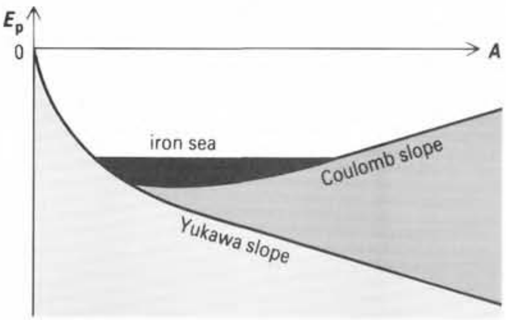



**Figure 9**
Geological section of the nuclear valley.

*C. Coulomb (1738–1806), a Frenchman who investigated the forces between electrically charged particles.

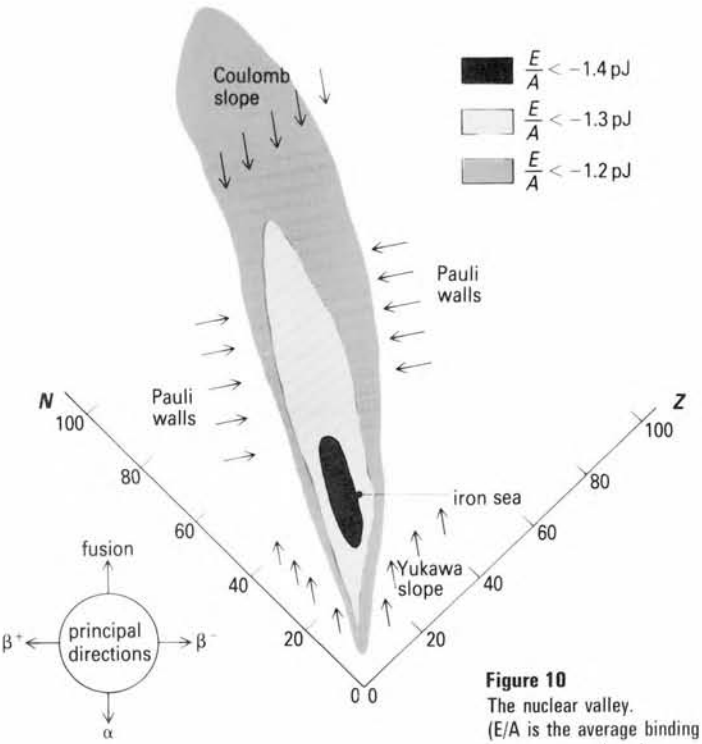



**Figure 10**
The nuclear valley.
(E/A is the average binding energy per nucleon.)

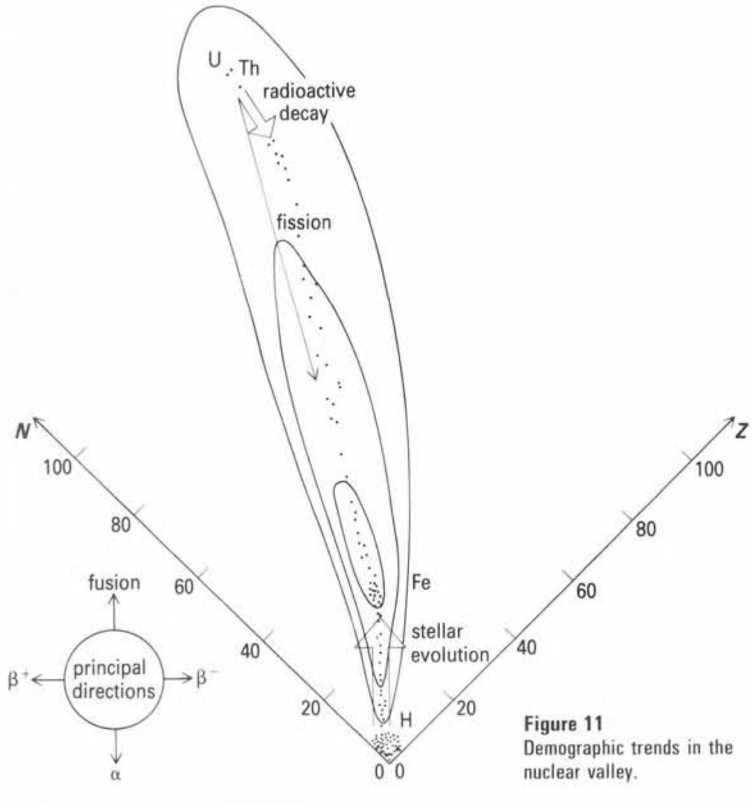



**Figure 11**
Demographic trends in the nuclear valley.

fraction of nuclear matter, like uranium and thorium, is still high on the Coulomb slope, leaking only very slowly to deeper places by radioactive decay. These elements feed our present nuclear power stations. When human society utilizes these nuclear differences, it makes use of a phenomenon as natural as a waterfall.

From the thermodynamic point of view there is no qualitative difference between hydroelectric and nuclear power. But in the case of nuclear power the thermodynamic potential difference is higher, by a factor of a million million! Any accident or misuse may be more dangerous in the same ratio. Nuclear literacy offers a chance to watch the wild beauty of this grand panorama. Both understanding the physics, and essential moral decisions, may help coming generations to choose a wise path in the nuclear canyon.

**General editor, Revised Nuffield Advanced Physics**
John Harris

**Consultant editor**
E. J. Wenham

**Editor of this book**
John Harris

**Contributors**
Paul Davies
Derek Eastham
Hester Greenstock
George Marx
Charles Taylor

Particles, imaging, and nuclei **is one of the background Readers for the Revised Nuffield Advanced Physics course. The five articles in this book extend concepts covered in the course and examine recent developments in physics. Paul Davies writes on the fundamental particles and forces of nature, Hester Greenstock writes on radioisotopes, Derek Eastham writes on the use of lasers in probing the atomic nucleus, Charles Taylor writes on imaging, and George Marx writes on the role of nuclear forces in the history of the Universe, and shows the relevance of nuclear physics to Man's energy needs.**

ISBN 0 582 35420 X

Longman