CROATICA CHEMICA ACTA
CCACAA **77** (1–2) 17–30 (2004)

ISSN-0011-1643
CCA-2897
Review

Shedding Light on Light*

John Murrell

Department of Chemistry, University of Sussex, Brighton, BN1 9QJ UK
(E-mail: J.N.Murrell@sussex.ac.uk)

RECEIVED NOVEMBER 11, 2002; ACCEPTED MARCH 18, 2003

Key words
velocity of light
waves and particles
colour
spectroscopy
Maxwell's equations
sunlight

A review is given of the history of the properties of light and the development of the concepts to help our understanding of these properties.

INTRODUCTION

There are four books describing early studies and views on light, optics and vision that are considered to be landmarks in this branch of science. The first of these is Euclid's *Optica*, published about 300BC, in which he postulated that light travelled in straight lines; he described the laws of reflection and described them mathematically. The second is by Ptolemy. He is best known for his 15 volume treatise on astronomy, *Almagest*, but he also wrote five books on optics around 100AD (only one of which survives to modern days). Ptolemy is credited with stressing the importance of basing theory on experiment, and in his *Optics* he studies colour, reflection and refraction, and mirrors of various shapes. The third is by Ibn al-Haitham, known later as Alhazen, who wrote a book on optics around 1000AD, which was later translated into Latin as *Opticae Thesaurus*. This describes experiments with mirrors and lenses, but is most famous for refuting a widely held view of Greek scholars (but not a unanimous view), that vision was due to rays emanating from the eyes, and hitting the object being viewed. He had several arguments against this view, the most telling being that it could not explain why some objects shone brightly (the Sun), and others dimly. He also noted the significance of the camera obscura, or pinhole camera, in which light from an object comes through a small hole and produces and inverted image. The last book, and the one which expresses views most consistent with modern science, is Newton's *Opticks*, of which more later.

THE VELOCITY OF LIGHT

The Greeks were the first to put in writing their views about the velocity of light. Aristotle believed that the speed of light was infinite, but he also quotes Empedocles (5BC) view that it was finite. »Empedocles says that light from the Sun arrives first in the intervening space before it comes to the eye, or reaches the Earth«. »…we should assume a time when the sun's ray was not yet seen, but was still travelling in the middle distance.« Over the next 1500 years both views were propagated, with probably the majority taking the line that it was infinite. Roger Bacon, the most important scientific thinker of his

---

* Dedicated to Professor Nenad Trinajstić on the occasion of his 65th birthday.

time, and the first person to give a geometrical description of optics, claimed in 1250 that the velocity of light was finite, but the distinguished astronomer Johann Kepler in 1600 said that it must be infinite because space could offer no resistance to its motion.

Galileo is reputed to have made the first serious attempt (in 1607) to decide the issue by experiment. Observers stationed on opposite sides of a valley had lamps with shutters. One observer opened his lamp, the other opened his when he first saw the flash, and the first observer tried to measure the time between his opening flash and the reply. However, light travels too fast to be determined by such a simple terrestrial experiment. It was however through Galileo's earlier astronomical observations of the moons of Jupiter that Ole Roemer deduced that the velocity was indeed finite. His work was made possible by some accurate tables that had been obtained in Paris by Giovanni Cassini for the movement of Jupiter's satellites. Cassini had noted that the eclipse times for Jupiter's closest moon Io varied as the distance between Jupiter and the Earth varied. Cassini is reputed to have first thought that this variation was due to the finite speed of light, but then rejected this view. Roemer took up the idea and in a paper presented to the Academie Royale des Science in 1667, which was summarised soon after in *Journal des Scavans*, there is the statement: »that for the distance of about 3000 leagues, such as is very near the bigness of the diameter of the Earth, light needs not one second of time«.*

The average time of Io's revolution is 42 h 28 min, but it appears to have a longer period of rotation when the Earth and Jupiter are moving apart, and a shorter period when they are moving together. Roemer concluded that the speed of light was such that it took 22 minutes to traverse the diameter of the earth's orbit, but as he had no accurate value for this distance, he gave no value for the speed in his paper. However, many others calculated a speed from his data, the first being Christiaan Huygens; after corresponding with Roemer, and eliciting more data, he deduced that light travelled 16.6 Earth diameters per second. Van Helden, writing in the *Journal for the History of Astronomy* in 1983, deduced that Roemer's observations and estimates of the Earth to Sun distance at that time would have given a speed of about 135 000 km s^{-1}. Later measurements of the eclipse times of Io by Delambre in 1790 and Glasenapp in 1874, together with more accurate values of the diameter of the earth's orbit, gave velocities around 300 000 km s^{-1}, which is quite close to modern values.

Roemer's view that the velocity of light was finite was not fully accepted until measurements of the so-called aberration of light were made by James Bradley in 1727. Bradley noted that the star Draconis apparently changed its position during the year in a manner that could not be due to parallax. He made further observations of other stars to show that the phenomenon was general, and arrived at the following explanation. Suppose that light from a star is coming down vertically to an observer on the Earth. If the velocity of light is finite, then because the earth is rotating about its axis, the telescope needs to be tilted away from the vertical in order that the starlight can pass down the length of the telescope. The angle to which it must be tilted depends on the ratio of the rotational velocity of the Earth and the velocity of light; the tangent of this angle is equal to this ratio. By following individual stars as the Earth rotates about the sun Bradley saw that their positions followed elliptic paths, from which he was able to calculate the aberration angle to be 20.2 min, and the tangent to be 9.8×10^{-5}. Knowing that the rotational velocity at his observatory was about 30 km s^{-1}, Bradley deduced that the velocity of light is about 300 000 km s^{-1}.

To improve on Galileo's attempt to measure the velocity of light by measurements on the earth's surface one needs to have a fast frequency modulator that can be interposed between transmitted and returned light beams. The first people to succeed with this were Armand Fizeau and Jean Foucault. Fizeau in 1849 used a rotating toothed wheel to chop a light beam, which then passed to a reflector at a distance of 8.633 km before returning to the wheel. The first interruption of the light occurred when the wheel rotated at 12.6 turns per second, and with 720 teeth in the wheel this gave a time for the light to pass 17.2 km as $(1/12.2 \times 1440) = 5.69 \times 10^{-5}$ s, giving a velocity of light of 313 000 km s^{-1}.

Foucault's experiment was reported in 1862. He used a rapidly rotating mirror to deflect the returning light beam. A rotation speed of 500 turns per second and a path length of 20 m deflected light to the observer by 0.7 mm, and this led to a velocity of light of 298 000 ± 500 km s^{-1}. Foucault's approach was later used in an improved form by Newcombe and Michelson, which finally led in 1882 to Michelson's result of 299 853 ± 60 km s^{-1}.

Mechanical modulating devices have an upper limit to the frequency at which they can be operated; typically of the order of kHz. This gives a lower limit to the time interval that can be measured and hence path lengths of several kilometres have to be used to determine an accurate value for the velocity. Much higher frequencies can be obtained by electro-optical modulation, and such a device was first used by Karolus and Mittaelstaed in

---

* An English translation of this report is in the *Philosophical Transactions of the Royal Society of London* (vol 12, 893 (1677–1678), from which the quotation is taken.

1928. Their device used polarised light produced by a Nicol prism (a calcite prism cut at certain angles that lets through only one plane of polarisation), which was then modulated by a Kerr cell before passing through another Nicol prism which is rotated by 90° to cut out the other plane of polarisation. A Kerr cell is a device that converts plane polarised light to elliptically polarised light by applying an electric field to an isotropic material. Early cells used a liquid such as nitrobenzene, and needed large electric fields to be effective, but the technology has now been advanced by the use of certain solid materials and rather small fields. The transmitted light is then reflected from a distant mirror and analysed through a Nicol prism and Kerr cell operating in phase with the first. Modulation frequencies of about 10 MHz were achieved in this way, and path lengths of the order of 100 m could be used. Such devices were later modified for use by others, notably Bergstrand, and when the velocity of light was standardised they became the basis for constructing very accurate geodetic (surveying) instruments.

More recent methods of measuring the velocity of light have largely been based on accurate methods for determining the wavelength and the frequency, the velocity being equal to the product of the two. Essen and Gordon-Smith (1945) were the first to use this method by establishing standing microwaves in a cavity resonator; standing waves are established if the length is an integer number of half wavelengths. A fixed frequency was established by a quartz crystal oscillator driven at 100 kHz, which was frequency multiplied to the region of 3000 MHz, and known to an accuracy of about 2 in 10^6. Later developments by Essen (1950) using this approach gave the velocity of light as 299 792.5 ± 3 km s^{-1}. This was outside the limits of a statistically recommended value proposed by Birge in 1941, but it was later proved to be a better value. Following the development of radar techniques in the 1940s the timing of pulsed radio waves from an aircraft to a ground station and back (Aslakson 1947), also led to doubts about the Birge value.

Wavelengths can be determined with great accuracy by an interferometer. In this a beam of monochromatic light is split in two, and after passing along slightly different path lengths the two beams are recombined before passing to a detector. If the path lengths differ by an integer number of wavelengths light from the two beams will enhance each other, if they differ by a half wavelength they will exactly cancel. An optical interferometer is usually constructed so that regions of constructive and destructive interference appear as bands or rings on the detector, and from the distance between these the wavelength can be deduced. A Michelson interferometer operates on light reflected in two mirrors, one fixed and one movable, and a Fabry-Perot interferometer uses the fringes produced by multiple reflections in plates of constant thickness.

It is not easy to produce monochromatic visible light whose frequency is known with high accuracy, so that interferometric techniques for measuring the velocity of light have used either microwave or radio-wave radiation. Froome in 1952 used a stabilised klystron to provide microwaves of 24.005 GHz. The interferometer was essentially of the Michelson type, with one beam going from a horn to a reflecting and movable mirror, and the other passing through a fixed wave guide. Both beams were combined at a superheterodyne receiver. Froome's result for the velocity of light agreed with that of Essen with an uncertainty of 0.1 km s^{-1}.

Other methods for determining the velocity of light have been based on the ratio of magnetic to electrical units, which are connected through Maxwell's equations as will be explained later.

WAVES OR PARTICLES

That light travels from a source to a detector (the eye), and that light transmits energy, was clear from early times, and from the 17th century there was speculation about whether its properties could be explained by particles (corpuscles), or by waves. The particle model was favoured by the fact that light travels in straight lines (shadows are sharp if the light comes from a point source), and that on reflection the angle of incidence is equal to the angle of reflection. Against the particle model is the fact that crossed light beams show no evidence of particle collisions.

The law describing the refraction of light on passing from one medium to another was discovered by Snell around 1621; the ratio of the sine of the angle of incidence to the sine of the angle of refraction is constant for a given interface between two media.* This observation, together with the law of reflection, was explained by Descartes in 1637 in his book *La Dioptrique*, using the particle model.** Reflection followed the mechanics of

---

* Snell discovered his law by experiment and communicated his result to several people by manuscript. It is said that he actually expressed his law in terms of cosecants, and it was Descartes who replaced this by sines.

** The word particle presents an oversimplification of Descartes ideas. He supposed that the vehicle of light was what he called matter of the second kind composed of closely packed assemblages of globules whose size is intermediate between that of vortex matter and ponderable matter. Vision is compared with the perception of the presence of objects which a blind man obtains by the use of his stick; the transmission of pressure along the stick from the object to the hand being analogous to the transmission of pressure from a luminous object to the eye by the 'second kind' of matter.

particles hitting an elastic surface, and refraction was due to the vertical component of the particle motion being slowed down by the surface when passing from a dense to a less dense medium (the horizontal motion being unchanged). Descartes argued that particles travelled faster in dense than in light media, on the grounds that a ball rolled faster on a hard table than across a soft carpet; one could argue the contrary, and indeed this was done by Fermat after he had read *La Dioptrique*. He made the alternative proposal that in refraction the light took a path that took least time; in the more dense medium the path was closer to the normal because the light moved more slowly.

Experimental confirmation that the velocity of light was less in the more dense medium was confirmed much later by the experiments of Fizeau and Foucault in the second half of the 19th century. However, it was later shown that Fermat's principle did not apply for light hitting curved surfaces, and the more general law which we now associate with his name is that the optimum path differs from its near neighbours only in the square of small quantities (the path can be a minimum or a maximum in the time taken).

A wave theory of light was first suggested by Francesco Grimaldi in his book *Physico-Mathesis de Lumine* published in 1665. This followed his observation that shadows were not precisely sharp so that light did not follow exactly straight lines. He also noted that the edges of shadows were often coloured, and concluded that light was a fluid capable of wave like motion, different colours being associated with different frequencies. He introduced the word diffraction (breaking-up) to describe this phenomenon. Hooke, in his *Micrographia*, published in 1667, also advanced the idea that light was a small and rapid vibratory motion of the medium, and he attributed the colours of thin plates or films to the interference of light waves. Hooke had earlier studied with Boyle, who had also noted these colours. Hooke specifically examines the question of how light is transmitted through space. He gives five requirements for this, the last of which is that »this motion is propagated every way with equal velocity, whence necessarily every pulse or vibrations of the luminous body will generate a sphere, which will continually increase and grow bigger just after the same manner (although infinitely swifter) as the waves or rings on the surface of the water do swell into bigger and bigger circles about a point of it, where by the sinking of a stone the motion was begun«. Hooke went on to use this model to explain the phenomena of refraction and of colour; his views on the latter were completely overthrown by the experiments made later by Newton.

Huygens, in his publication *Traite de la lumiere* (1690, but there was an early version in 1678), greatly improved the wave model, and showed that it could explain reflection and refraction, and also the phenomenon of double refraction, which had been discovered by Erasmus Bartholine in 1670. When light passes through certain crystals like Iceland Spar (calcite) it is split into two beams, one of which (the ordinary ray) follows Snell's law of refraction, and the other (the extraordinary ray) does not. Huygens showed by experiment that the two rays differed from ordinary light in that when passed through a second crystal they were either split into two or un-split depending on the orientation of the second crystal.

Huygens' most important idea, which is accepted today, was that every surface element at the front of a wave can be regarded as a source of secondary wave motion which will be propagated outwards in the form of a sphere, and the wave front at a later time is the envelope of the secondary waves. Huygen's interpretation of reflection and refraction is based on what happens to the wave front when it hits the surface.

The problem for the wave model was that it was assumed to require a transmitting medium. (it had already been shown by Torricelli that light is transmitted as easily in a vacuum as it is in air), and Huygens took this to be a luminferous aether. The idea that action at a distance could only occur through a medium called aether goes back to the Greeks, but it was Descartes and his followers who were the first to give it scientific importance by ascribing mechanical properties to it. Huygens assumed that light waves were longitudinal, as are sound waves and only much later was it was shown that they are transverse.

The most famous early supporter of the particle model was Newton, and this resulted in one of several disagreements with Hooke. Following Newton's first scientific publication in 1672 on the coloured components of white light, Hooke charged him with holding the doctrine that light is a material substance.* Newton in reply said that he made no such hypothesis, and to link his results to Hookes ideas he inferred that corpuscles corresponding to different colours would excite vibrations of different types in the aether.

In a paper sent to the Royal Society in 1675 Newton assumed that the rays of light are »small bodies emitted every way from shining substances«, and »where they impinge on any refracting or reflecting superficies, must as necessarily excite vibrations in the aether, as stones do in water when they are thrown into it.« His support

---

* Hooke is reported to have said about Newton's paper that what was original in the paper was wrong, and what was correct was stolen from him.

for the particle model is, here, unquestionable. However, in his famous book *Opticks*, first published in 1704, he relates the vibrations of light and sound in the following statement. »Do not several sorts of rays make vibrations of several bignesses, which according to their bignesses excite sensations of several colours, much after the manner that the vibrations of the air, according to their several bignesses, excite sensations of several sounds? And particularly do not the most refrangible rays excite the shortest vibrations for making a sensation of deep violet, the least refrangible the largest for making a sensation of deep red, and the several intermediate sorts of rays vibrations of several intermediate bignesses to make the sensations of the intermediate colours.« If one takes »shortest« and »longest« to refer to wavelength, then we have a correct association of colour with the periodicity of the light waves.

Newton retained his corpuscular model for topics where the wave model seems more appropriate. He repeated and extended Grimaldi's experiments on shadows and concluded that bodies act upon light at a distance, thereby bending rays. He also used the model to explain the colours of thin films, and what we now call Newton's rings; these are the interference patterns observed from reflected light which has been passed through a convex lens sitting on a piece of plane glass or another convex piece. He supposed that »every ray of light in its passage through any refracting surface is put into a certain transient constitution or state which in the progress of the ray returns at equal intervals and disposes of the ray at every return to be easily transmitted through the next refracting surface, and between the returns to be easily reflected by it«. The length of fit for easy transmission he assumed to depend on the colour of the light.

From the phenomenon of double refraction Newton concluded that a ray obtained by double refraction differed from ordinary light in the same way that a rod of rectangular cross section differs from one of circular cross section. He wrote: »That every ray of light has therefore two opposite sides, originally endued with a property on which the unusual refraction depends, and the other two opposite sides not endued with that property«. He concluded that the refraction from a crystal depends on the relation of these sides to the planes of the crystal. In this analysis of double refraction he came close to the idea that light can be split into two beams which differ in their planes of polarisation. Newton knew that Huygens had explained double refraction with his wave model, but noted that he was at a loss to explain the behaviour of these waves when passed through a second crystal. To Newton the phenomena were »inexplicable if light be nothing else than pression or motion propagated through aether«.

Newton was considered to be the champion of the particle model largely by his scientific supporters who came later, and the model dominated thinking for the next century. It did not lose its dominance until the work of Thomas Young. In his first paper on optics, published in 1800, Young cast doubt on the particle model on the basis that the velocity of light was the same whether the light came from a violent source like the sun, or from a feeble spark, and in 1802 he explained Newton's rings within the wave model. He drew on the observation of water waves saying that if waves pass through two channels with different speeds then if the elevations of one coincide with the depressions of the other, the surface of the water will remain smooth. »Now I maintain that similar effects take place whenever two portions of light are thus mixed; and this I call the general law of the interference of light.«

In a lecture to the Royal Society of London in 1803 (published in *Philosophical Transactions* in the following year) Young described the double slit experiment, and concluded that: »It will not be denied by the most prejudiced that the fringes are produced by the interference of two portions of light«. His first experiments were actually done by splitting a narrow light beam with a thin card held edgewise to the beam and observing the light and dark bands (fringes) that occur when the two beams pass on to overlap and interfere. In later experiments the beam was passed through two slits slightly separated from one another. From the separation of the fringes he was able to show that the wavelength of light was about a millionth of a metre, and from this small value he was able to explain why light travels around objects in almost straight lines. Young's first experiments were on white light, which, having a mixture of wavelengths, does not give sharp fringes, but he later used coloured light and determined the wavelengths of each colour.

The next step in the support of the wave model came from the observation of Malus (1808) that light reflected from a glass surface was polarised, and Arago deduced that the same was true of the refracted ray. In 1815 Brewster showed that when the reflected and refracted rays were at right angles then polarisation was complete; this could be confirmed by passing the resulting rays through Iceland Spar when one and not two images were seen. These observations seemed to support Newton's idea that light particles possessed sides, and Young was at first unable to fit them to a wave model. Particularly puzzling were the observations of Arago and Fresnel that two oppositely polarised rays did not interfere with one another. Then in 1817 Young suggested in a letter to Arago that light vibrations occurred transversely, like water waves, and not longitudinally like sound waves. From this the phenomenon of polarisation could be understood, because there would be two possible modes of vibration at right angles to the direction of the beam, and these could be separated, when the light met a surface. Young's view was soon adopted by others, notably by Fresnel who showed that it could explain all the known phenomena of optics, particularly the non interference of oppositely polarised rays. However, he also noted that although longitudinal waves could be transmitted through a gas-like medium, this was not true of transverse waves, the luminiferous aether therefore needed to possess some solid-like properties. These difficulties were not resolved until Maxwell's electromagnetic wave theory was developed.

UNDERSTANDING COLOUR

The Greeks, of course had ideas about colour, but these were mainly concerned with linking colours to other concepts; the colours of earth, fire, air and water; the colours of harmonic notes, *etc*. Aristotle believed that blue and yellow were primary colours, and he noted the effect of mixing colours, and that these primary colours produced green. The first scientific explanations of colour were given in the 17th century. Hooke, in his book Micrographia (1667), connected colour with the refraction of light; the greater the refraction the more the colour moved towards the violet end of the spectrum. Newton's predecessor in the Lucasian chair of mathematics and astronomy at Cambridge was Isaac Barrow. Newton is said to have assisted him in preparing lectures on optics, and these lectures were later published (1669). Barrow explained colour in phrases like 'red is that which sends out light more concentrated than usual but broken up by dark interstices, and yellow consists mainly of white with some red interspersed'. These ideas were soon to be overturned by Newton's experiments.

Newton's first researches on light were carried out in 1666, but were not published until 1672 (in the *Philosophical Transactions of the Royal Society*). This paper contains his famous description of the splitting of white light into the rainbow colours by a prism. »I procured me a triangular glass prism, to try therewith the celebrated phenomenon of colours. And in order thereto having darkened my chamber, and made a small hole in the window-shuts, to let in a convenient quantity of the sun's light, I placed my prism at his entrance, that it might be thereby refracted to the opposite wall. It was at first a very pleasing divertissement, to view the vivid and intense colours produced thereby; but after a while applying myself to consider them more circumspectly, I became surprised to see them in an oblong form; which according to the received laws of refraction, I expected should have been circular.«

Newton went on to show that white light is a mixture of rays of every variety of colour, and the elongation of the spectrum is due to the differences in the refractive power of glass for these rays. The key observations of Newton were to show that one colour was never split up into other colours, no matter how much it was later refracted, and to show that the rainbow colours when recombined gave back white light. He concluded that »if the sun's light consisted of but one sort of rays, there would be but one colour in the whole world, nor would it be possible to produce any new colour by reflections or refractions, and by consequence that the variety of colours depends on the composition of light.«

One important consequence of Newton's observation that the refraction angle depends on colour, is to provide an understanding of the fact that telescopes based on lenses with spherical curvature cannot give sharp images because of so-called chromatic aberration, Newton invented a reflecting telescope to overcome this problem, and this telescope was already in the possession of the Royal Society when he submitted his 1672 paper.

In 1692 Newton left a light burning in his room when he went to chapel, which destroyed his papers, among them a large work on optics, containing the experiments and researches of 20 years. The first edition of *Opticks* was not published until 1704, and the last edition, corrected by the author's own hand, and left before his death with his bookseller, was published in 1730. Newton's *Opticks* contains some interesting speculations about the nature of light, which he had not been able to prove, some of which would later be found to be true or have some elements of truth in them. Here is a selection:

Query 1. Do not bodies act upon light at a distance, and by their action bend its rays; and is not this action strongest at the least distance?

Query 5. Do not bodies and light act mutually upon one another; that is to say, bodies upon light in emitting, reflecting, refracting, and inflecting it, and light upon bodies for heating them and putting their parts into a vibrating motion wherein heat consists?

Query 6. Do not black bodies conserve heat more easily from light than those of other colours do, by reason that the light falling on them is not reflected outwards, but enters the bodies, and is often reflected and refracted within them, until it be stifled and lost?

Query 8. Do not all fixed bodies when heated beyond a certain degree, emit light and shine; and is not this emission performed by the vibrating motions of their parts?

Query 29. Are not the rays of light very small bodies emitted from shining substances?

Query 30. Are not gross bodies and light convertible into one another, and may not bodies receive much of their activity from the particles of light that enter their composition?

In 1800, William Herschel the astronomer carried out experiments to see how much heat was associated with the different parts of the sun's spectrum. He placed thermometers with blackened bulbs in different regions of the spectrum and had control thermometers outside these regions. He noticed that the temperature rose from the violet to the red end of the spectrum, and that the control thermometer beyond the red showed the highest temperature of all. Herschel called these rays calorific rays, later to be called infrared. Humans detect infrared only through the heating effect it produces, but some snakes and insects do have infrared detectors. In 1847 Fizeau and Foucault showed that infrared radiation could be reflected and refracted, and that it could form interference patterns; these observations established that it was of the same type of radiation as visible light.

Soon after Hershel's discovery, in 1801, Johann Rittner showed that there was radiation beyond the violet region. He did this by showing that silver chloride coated paper was blackened beyond the violet end of the spectrum. He called the rays, chemical rays; later to be renamed ultraviolet. Other regions of the electromagnetic spectrum were discovered after the implications of Maxwell's equations were appreciated.

Newton guessed that colour is a manifestation of the interaction of light with the eye; the retina passing signals to the brain. Query 12 of *Opticks* says: »Do not the rays of light in falling upon the bottom of the eye excite vibrations in the *Tunica Retina?* Which vibrations, being propagated along the solid fibres of the optick nerves into the brain cause the sensation of seeing«, and in Query 13: »Do not several sorts of rays make vibrations of several bignesses, which according to their bignesses excite sensations of several colours«.

Speculation on how the retina reacted to light led George Palmer in 1786 to propose that colour is based on maximal sensitivity of retinal particles. It was reported in 1781 that Giros von Gentilly held that there were three types of molecule or membrane in the retina corresponding to three kinds of light, but it was not until 1802 that the idea of only three receptors was proposed more forcefully by Thomas Young. »As it is almost impossible to conceive each sensitive part of the retina to contain an infinite number of particles, each vibrating in perfect unison with every possible undulation, it becomes necessary to suppose that the number is limited; for instance to the three principal colours red, yellow and blue«.* Young carried out experiments on colour mixing by overlapping his three primary colours.

Dalton (1798), had earlier described his own colour blindness (daltonism), but proposed that his deficiency was due to the fact that the vitreous humour in his eyes was tinted blue. This was found to be false when his eyes were examined on his death in 1844, and in recent times it has been shown from DNA extracted from Dalton's preserved eyes that he had a genetic deficiency, the loss of one receptor making him a deuteranope.

Colour mixing was studied extensively by Helmholtz, and although it is said that he believed that there must be at least five receptors, the trichromatic theory of colour is usually called the Young-Helmholtz theory. Ewald-Herring examined the occurrence of after images (red after green, for example), and proposed that visual signals occurred in opposing pairs (yellow-blue, red-green, white-black)

The science of vision is a subject that can only be given a brief comment in this article. The receptors were first seen by microscopy in the 17th century, and by the 19th century it was known that there were two types, called, from their shapes, rods and cones. Rods are responsible for black and white images in dull light, and the cones are responsible for colour in bright light. The retina contains about a hundred million rods and three million cones. In 1851 H. Muller extracted a compound from rods, which was called visual purple, and in 1876 Boll discovered that this was bleached by light. In 1878 Kuehne isolated rhodopsin, and this was later shown to be made up of two components, a protein called opsin, and a compound 11-*cis* retinal, which has a side chain of six conjugated carbon-carbon double bonds. When retinal absorbs light it undergoes isomerisation to an all-*trans* form, and this leads to a chain of events ending in the closure of ion channels and the sending of an electrical signal to the brain. In the 1940s it was shown that this process is extremely efficient; a rod absorbing a single photon leads to isomerisation of the retinal in about 50 % of cases. The process is also very fast, occurring in about 200 fs (200×10^{-15} s)). The same chromophore, retinal is also present in the cones, but the proteins (opsins) that it is linked to are different from that in the rods, and each of the three colour detectors have different opsins whose effect is to move the absorption region of retinal to different wavelengths.

SPECTROSCOPY

In 1752 Thomas Melvill reported that the yellow colour that appears when common salt is placed in a flame had a definite degree of refrangibility. Later, in 1802, Wollaston noted that the spectrum of sunlight was crossed by several dark lines, and in 1813 Fraunhofer measured the wavelengths of these lines by using a diffraction grating. One of these lines, called D by Fraunhofer, was found to have the same wavelength as that of the yellow colour from common salt. In 1826 Talbot examined the light from various flames with a prism and

---

* It is said that Michael Lomonosov (the father of Russian science), proposed the three receptor basis of colour vision much earlier, in 1757.

noted that the spectrum indicated chemical substances, which would otherwise require laborious chemical analysis to detect; he was able to distinguish between the red colours of lithium and strontium salts. With such observations spectroscopy was born as a subject of both pure and applied research.

The connection between the wavelengths of dark lines in the sun's spectrum and the bright spectral lines from flames was picked up by several people, and Ångström in particular reported in 1853 that an incandescent gas emits luminous rays of the same refrangibility as those which it absorbs at ordinary temperatures; the connection between absorption and emission of light was being made, and Stokes drew an analogy with a stretched string which if struck would give off a note and which would take up the same note if it was sounded in air.

In the second half of the 19th century a great deal of information was being obtained on the line spectra of the elements, and in 1871 Johnstone Stoney proposed that the best way to organise these was on the basis of their wave number, which is the reciprocal of the wave length. This revealed common patterns in the intervals of multiplet spectra. In 1885 Balmer produced the first simple formula for the wavelengths of the four lines which were known for the hydrogen spectrum. Similar patterns were then investigated for other simple spectra, such as those for the alkali metals, and in 1890 Rydberg gave a generalisation of the Balmer formula for the wave numbers of the hydrogen and alkali spectra:

$$n = n_0 + R/(m + p)^2 \qquad (1)$$

where R is a common constant for all spectral series (the Rydberg constant), n_0 is a constant of the particular series, and m is an integer. Rydberg and Ritz then noted that the wave numbers of several series for the same element could be represented by differences of two Rydberg type formulas, and in 1908 Ritz put forward the combination principle that to the spectrum of any element there corresponds a set of terms such that the wave numbers of the lines of a spectrum are the differences of these terms taken in pairs.

The relationship between spectral wave numbers and energy levels was not seen until Planck began quantum theory in 1900, and Rutherford proposed his model of the atom in 1911. Nicholson, also in 1911, then recognised that the production of atomic spectra was essentially a quantum phenomenon, and he proposed that in the production of a spectral line the angular momentum could only rise or fall by discrete amounts. Nernst in the same year said that it was a necessary consequence of quantum theory that the vibrational and rotational motions of molecules must vary discontinuously.

The first explanation of a spectrum in what we would now call old or semi-classical quantum theory was made by Bjerrum in 1912. He attributed the infrared spectrum of gaseous hydrogen chloride to a vibration of the positive hydrogen and negative chlorine atoms against one another, and the pattern of lines within this spectrum to rotations of the molecule with the rotational energies being multiples of $h\nu$, where ν is the number of revolutions per second. The energies of rotation were then given by the classical formula involving the moment of inertia (I):

$$0.5\, I(2\pi\nu)^2 = nh\nu \quad \text{where} \quad n = 0,1,2,\ldots. \qquad (2)$$

The explanation of atomic spectral lines by old quantum theory was given by Bohr in 1913. He stated more clearly than Nicholson that two distinct energy states of the atoms are associated with each spectral line, and the fact that the wave numbers of the hydrogen spectral series were reproduced exactly, using only the values of the fundamental constants (h, c, e, and m) was at the time a stunning verification of the theory. It was only later when the theory was seen to be inadequate for reproducing the wave numbers of more complicated atoms that it was eventually replaced by Schrödinger's wave mechanics.

The equilibrium between radiation and matter was developed in quantum theory in 1916 by Einstein. He was able to give a simple proof of Planck's black-body radiation law, but of equal importance was his conclusion that a necessary consequence of the equilibrium was that there were two processes for emission, one stimulated, which could be considered as the inverse of the absorption law, and the other was spontaneous. The relative probabilities of these two forms depend on the frequency as well as the intensity of the radiation field, and the importance of stimulated emission was not realised until much later with the development of lasers.

MAXWELL'S EQUATIONS

Until the early 19th century electricity and magnetism were considered as separate subjects. Both static electricity and lodestone, the naturally occurring magnet, were known to ancient civilisations in both East and West. It was also known that lodestone could be used to magnetise iron. From his studies on lightening Franklin suggested that the charges associated with static electricity could be mobile and were conserved, and the first continuous source of electricity was produced by Volta at the end of the 18th century; the voltaic cell having two different metals dipped in acid was produced in 1799. The best steady currents were produced through a discovery of Seebeck, who showed in 1822 that the junction between two different metals produced an electric potential, and a current when the circuit was closed. With this source of electricity Ohm was able to establish the relations between potential, current and resistance.

The first connection between magnetism and electricity was made by Oersted in 1820 when he showed that an electric current when passed through a wire influenced the orientation of a near-bye compass needle, and the force on the magnet had the effect of driving it in a circle around the wire. The report of this experiment had immediate worldwide recognition. Henry used the knowledge to manufacture an electromagnet that could lift a ton of iron, and Ampere found that two parallel wires carrying current attracted each other if the currents were flowing in the same direction, and repelled each other if the currents flowed in opposite directions. An important conclusion of Ampere was that the magnetic forces arising directly from magnets were exactly the same as the magnetic forces produced by electric currents. Hertz later showed that there was a similar unity of electric forces; those produced from electric charges were the same as those produced by moving magnetic fields.

The next steps were taken by Michael Faraday who was working with Humphrey Davy at the Royal Institution in London. He showed in 1821 that when current was passed through a rod that was free to move, and one end of the rod was near a magnet, then the tip of the rod circled the magnet; this was the first electric motor. In 1831 Faraday showed that when a magnet moved through a coil of wire it induced a current (the first dynamo), and in the same year, that a current could be induced in a coil of wire by passing a current through a neighbouring coil (the first induction coil and transformer). To understand these experiments, and particularly the fact that they involved actions over distance (like gravity), Faraday introduced the concept of fields, which would impose forces on objects that they covered. An electric field produces a force on an electrically charged particle, and a magnetic field produces a force on a magnetic particle. The important feature of Faraday's experiments was in showing that a changing magnetic field produces an electric field, and a changing electric field produces a magnetic field.

Although the concept of fields was slow to be adopted by physicists, there were some who expressed Faraday's results in mathematical form. Lord Kelvin in doing this became convinced that not only was there a link between magnetism and electricity, but that light was connected to both. Faraday showed in 1845 that this was indeed likely, by the observation that when polarised light was passed through a magnetic field the plane of polarisation was rotated. He asked »Whether it was not possible that the vibrations which in a certain theory are assumed to account for radiation and radiant phenomena may not occur along the lines of force which connect particles and consequently masses of matter together«.

It had been shown by Fizeau in 1850 that electricity travelled along copper wires at approximately two thirds the speed of light, and the link between electricity, magnetism, and light, was made even stronger by Kohlrausch and Weber in 1856. Coulomb had earlier (1785) shown that the force between electrostatic charges obeyed an inverse square law, like that for gravity, and that if a long bar magnet was taken to be made up of two separated magnetic poles, then the force between magnetic poles also obeyed the inverse square law. The electrostatic system based on the first of these laws described quantities in a set of units known as electrostatic units (esu), the magnetic system led to another set of units called electromagnetic (eventually cgs), and because of the experiments of Faraday and others it was possible to connect the two; what was needed was to measure a quantity such as electric charge by two different methods and obtain its value in both esu and cgs units. Kohlrausch and Weber measured the charge on a capacitor using a Leydon Jar (in esu), and the charge was then measured (in cgs) by discharging it through a galvanometer. The ratio of the two has the dimensions of velocity, and the value obtained was 310 740 km s^{-1}, which was very close to the value then accepted for the velocity of light. Improvements on such measurements were made later, and by 1907 Rosa and Dorsey had obtained a value for the velocity of light by this method of 299 788 ± 30 km s^{-1}.

Maxwell developed his equations for electromagnetic fields in a series of papers between 1861 and 1868. Two of Maxwell's equations give a description of Coulomb's law for electric fields and for magnetic fields (although there is an important difference between the equations because magnetic poles do not exist in isolation). A third equation is Ampere's law relating the magnetic field vector and the current density, and the fourth describes Faraday's law of induction. Maxwell saw that Ampere's law needed modification if the charge density varied with time, and he added a term to that equation that he called the displacement current. The resulting four equations are one of the foundations of classical physics, and when tested for invariance under moving coordinate systems at constant velocity they led to the equations of special relativity.

The relevance of Maxwell's equations to light comes when one examines their form in space which is empty of anything except electric and magnetic fields. From the four equations one can then prove that both the electric and magnetic fields must satisfy the conventional equation for harmonic waves; for the electric field this can be written

$$\nabla^2 \boldsymbol{E} = \mu_0 \, \varepsilon_0 \, (\partial^2 \boldsymbol{E} / \partial t^2), \tag{3}$$

where μ_0 and ε_0 are magnetic and electric parameters of free space respectively. A similar equation holds for the magnetic field $\boldsymbol{B}$. The factor $(\mu_0 \varepsilon_0)^{-1/2}$ is the wave velocity, and Maxwell noted that its value was consistent with the known value for the velocity of light. This knowledge led to the realisation, by experiment and theory, of the whole electromagnetic spectrum.

Maxwell's equations made a huge impact on physicists, and several attempts were made to justify their validity, and particularly of Maxwell's introduction of the displacement current. The status of the equations was finally established by the experiments of Heinrich Hertz, who in 1887 showed that when there was an oscillating discharge of a condenser through a loop with a spark gap, a signal could be picked up, and a spark produced in a second loop that was some distance away. Unknown to Hertz a similar observation, and its implications, had been made by David Hughes some seven years earlier. Hughes showed that the signals from a spark could be picked up by a microphone fifty yards away. Sadly when his experiment was demonstrated before the president of the Royal Society and others, they were not impressed, and due to the discouraging opinions of his audience his paper did not appear until 1899. Hertz went on to establish that his electromagnetic signals could be reflected and refracted, and that their velocity was of the same order of magnitude as that of visible light; perhaps, most interestingly, that they passed through non-conductors.

Hertz made many studies of the transmission and receipt of his waves, and in 1888 he presented his results before an audience containing a young Italian called Marconi.

SUNLIGHT

The Sun does not give us heat, it gives us light which is then converted to heat when the light is absorbed on the Earth. The origin of the Sun's light is to be found in the more general question of the origin of starlight, and understanding this came quite late in the development of science.*

The first explanation of the Sun's energy was by Helmholtz (1867) who attributed it to gravity. In the condensation of interstellar matter to form a star the gravitational potential energy decreases, and one half of this energy is set free as kinetic energy, according to the virial theorem of mechanics. It is this kinetic energy that largely determines the temperature of the sun, which at its centre is thought to be about 15 million degrees. If we know the rate at which energy radiates from the sun then we can estimate the lifetime of this radiation, and 19th century physicists concluded that gravity provided enough energy for the sun to radiate for about 15 million years. However, there was another landmark in science in the 19th century, and that was Darwin's theory of evolution, and biologists argued that 15 million years was too short a time to allow the present diversity of life on Earth to be produced; the biologists were strongly supported by the geologists, but they were bitterly opposed by eminent physicists like Lord Kelvin.

The correct explanation of the Sun's energy is that it is due to nuclear reactions, but these were not known until Henri Becquerel's discovery of radioactivity in 1896. Becquerel showed that uranium salts, even when wrapped in paper, blackened a near bye photographic plate, and he also showed that the uranium rays imparted conductivity to gases, like the recently discovered X-rays of Röntgen. Pierre Curie and Laborde noted in 1903 that radium salts emitted heat and never cooled down, and in 1904 Rutherford said: »The discovery of the radioactive elements which in their disintegration liberate enormous amounts of energy, thus increases the possible limit of the duration of life on this planet and allows the time claimed by geologists and biologists for the process of evolution«. The age of the Earth shot up from millions to billions of years..

The combination of Einstein's mass-energy equation and radioactivity gave plenty of scope to explain the energy of stars. The most extreme view was taken by Eddington in the 1920s, who considered the consequences of assuming a complete conversion of mass to energy. However, no reaction has been found to completely annihilate nuclear matter. The biggest stumbling block was how the electrostatic repulsion between two positively charged nuclei could be overcome to release energy, but this was overcome when Gamow showed in 1928 that quantum mechanical tunnelling through potential barriers could provide an explanation. In the 1930s particle accelerators were built to study nuclear reactions, and the rates observed were found to be in line with this theory.

If nuclear reactions rely on the quantum mechanical tunnel effect, then the lightest nucleus, the proton, should be the most efficient. Von Weizsacker in 1938 concluded that a reaction between two protons, and another of protons with carbon, nitrogen and oxygen nuclei, which catalytically regenerates carbon (the so-called CNO cycle) were the most likely contenders for the Sun's Energy. Hans Bethe, in the same year, looked at the reactions of protons and other nuclei, going up the periodic table and reached the same conclusion; he also established that the proton-proton reaction involved helium in a later step. It is now clear that in the Sun the latter is much the most important, but the CNO reactions are important for heavier stars.

The dominant nuclear process in the Sun is that four protons combine to give a helium 4 nucleus, and as it had already been established by Aston in 1920 that this led to

---

* The history of the subject is described very well in Hans Bethe's Nobel lecture of 1967 from which this account is largely taken. This is accessible from the Internet.

a loss of mass of about 7 in a thousand, it was clear that it was this loss, converted to energy through the Einstein equation, that produces the Sun's energy. The most important steps in the proton-helium chain are as follows:

i. Two protons combine to give a deuterium nucleus, a positron and a neutrino.

ii. A deuteron and a proton combine to give helium 3, and radiation.

iii. Two helium 3 nuclei combine to give helium 4 and two protons.

The first of these processes is very slow, but there are enough protons in the Sun to produce the required rate of energy production; the two that follow are relatively rapid.

LIGHT AND RELATIVITY

The wave interpretation of light, which received such powerful support in the 19th century, was intimately tied to the concept of the aether. Young wrote: »I am disposed to believe that the luminiferous aether pervades the substance of all material bodies with little or no resistance, as freely perhaps as the wind passes through a grove of trees«. The big question was not so much what was the aether, but how was it affected by the movement of the earth.

The interpretation of aberration, for example, which was straightforward in the corpuscular model, was explained by Young, but only with the assumption that the aether did not move with the earth's motion; if the aether surrounding the earth was at rest, then the light waves would not partake of the motion of the telescope, and aberration would be observed. Many people famous in the field of optics set out to examine the implications of a moveable or stationary aether, but the crucial experiments of Michelson and Morley to test this in 1887 eventually led to the aether concept being abandoned.

The experiment was based on Michelson's development of the interferometer in which a monochromatic beam was split in two and later combined to give an interference pattern. To maximise sensitivity in the experiment a long path length was needed, and to eliminate the effect of varying temperature along the two paths, a length of about 11 m was achieved by multiple reflections over a short distance. Michelson pointed out that because the beams had to pass forwards and backwards along their paths, any difference due to the motion of the source was of second order in the source velocity (think of swimming forward and back against a current); the experiment was calculated to detect an effect due to a source velocity in the two paths of $3 \times 10^{-5}c$ (c being the velocity of light); which was to be interpreted as a difference in the earth and aether velocities.

The experiment was set up to detect a difference in the velocity of light between two beams one going with the direction of the earth's motion and one at right angles to this. Because the direction of motion was not known, the whole apparatus was designed to be rotated through 90° on the laboratory bench. Michelson and Morley expected to see changes in the interference fringes as the apparatus was rotated, but no such changes were observed. This result was confirmed by experiments at different times of the year, and later the experiment was repeated with other types of electromagnetic waves.

Many others did experiments with the same objective of Michelson and Morley; some also found a null result, and other positive results were later found to be unsound. There was one important interpretation of the experiment that had some truth, but that was not seen until much later. In 1892 Lodge mentioned in a paper in *Nature* that Fitzgerald had told him that the null result of Michelson and Morley could be explained if the dimensions of material bodies were slightly altered when they are in motion relative to the aether. This hypothesis was immediately adopted by Lorentz and was widely accepted by others, but it was not confirmed in several experiments; only later was it clear from the theory of Special Relativity that the Fitzgerald contraction in a moving system could not be measured by observers in that system.

How the Michaelson-Morley result led to Special Relativity is a story in itself. Several big names in physics contributed to the subject culminating in Lorentz's extension of Maxwell's equations to establish that the laws of physics would be the same in all bodies moving with uniform velocity; these equations were later to be called the Lorentz transformations. In the meantime, Poincare had said in lectures in 1899: »I regard it as very probable that optical phenomena depend only on the relative motions of the material bodies, luminous sources and optical apparatus concerned, and that this is true not merely as far as quantities of the order of the square of the aberration, but rigorously. By 1904 Poincare had defined a 'Principle of Relativity', in which the laws of physical phenomena must be the same for a fixed observer as for an observer who has a uniform motion of translation relative to him, and he also concluded that that no velocity could exceed the velocity of light.

That Einstein scooped the pot for Special Relativity is a very strange story. His position at the time is described in the next section, and even stranger is the fact that he did not get his Nobel Prize for Special or General Relativity. Einstein's papers are not renowned for his acknowledgement of other people's work, and his first paper on relativity is no exception. It has never been clear whether Einstein new of the Michelson-Morley result; he certainly new of the work of Lorentz that followed it, but probably not of Poincare's contributions. Einstein based his theory on the assumption that the velocity of light was independent of the velocity of the source (which was one way to interpret the Michelson-Morley result). The famous equation $E = mc^2$ followed from his theory, but this equation had been suggested earlier by others. This article is not the place to give either the theory or history of Special Relativity, nor its development to General Relativity in 1915 for accelerating systems. The latter included a description of how light was influenced by a gravitational field (a proper treatment of Newton's query on this matter). There are several excellent biographies of Einstein that cover these matters, and the book by Abraham Pais is highly recommended.

EINSTEIN'S NOBEL PRIZE

In the summer of 1900 four students obtained diplomas at the technical high school in Zürich (the ETH), which qualified them to teach high school mathematics and physics. Three of them were immediately appointed as assistants at the ETH, which would provide a route for them to study for a PhD. One of them, Albert Einstein, was refused such a position. Einstein was certainly not a conscientious student, being accustomed to miss classes. His physics professor, Heinrich Weber was reputed to have said something like 'You are a smart boy Einstein, a very smart boy, but you have one great fault: you do not let yourself be told anything'. He was probably too harsh as Einstein's Nobel Prize paper dealt with a topic of Weber's own research, on which he had lectured to Einstein in 1899.

Einstein tried hard to find a position at other universities but failed, and after short spells in two high schools he obtained, through the influence of the father of his friend and fellow student Marcel Grossman, the post of technical expert third class at the federal patent office in Berne. The rest, as they say, is history.

Einstein had his first scientific paper published in 1901; he used thermodynamic arguments in an attempt to deduce a universal function for intermolecular forces. A further paper on similar lines was published in 1902, and then three papers on the foundations of statistical mechanics, the last being in 1904. He was then still at the patent office (his appointment had just been made permanent), but he had not yet obtained his doctorate. None of his published papers would have aroused much interest in the scientific community, although looking at them in retrospect one can see many influences on his later work. He completed a thesis for his doctorate in April 1905, with the title *On a new determination of molecular dimensions*, and submitted it to the University of Zürich in July, where it was quickly accepted for the degree. The delay in submission was probably due to the fact that he was completing two landmark papers, one on Brownian motion (sent to the journal in May), and the other on Special Relativity (sent in June); a second paper on Special Relativity was submitted in September and contains the famous equation $E = mc^2$.

The paper that was to lead to Einstein's Nobel Prize was submitted prior to his PhD, in March 1905; its title was *On a heuristic point of view concerning the generation and conversion of light*. There have been many comments on Einstein's use of the word heuristic. Literally it means 'relating to discovery', but it is generally felt that Einstein's use implies 'based on evidence but being rather speculative'. The paper contains the following statement: 'monochromatic radiation of low density behaves in thermodynamic respect as if it consists of mutually independent energy quanta of magnitude $h\nu$'.

Einstein clearly considered the proposal in this paper as a hypothesis, and he later said that it was the only one of his 1905 papers that was revolutionary. Others thought the same, and it was several years before the quantisation of light was widely accepted by the physics community. For example, in 1913 Einstein was proposed for membership of the Prussian Academy, and two of his sponsors were famous exponents of energy quantisation, Planck and Nernst. Their support was immensely strong, but they also made the following statement 'That he may sometimes have missed the target in his speculations, as, for example, in his hypothesis of light quanta, cannot really be held too much against him, for it is not possible to introduce really new ideas even in the most exact sciences without sometimes taking a risk'. Einstein was not awarded the Nobel Prize for light quantisation until 1922.

The paper starts with the sentence »A profound formal difference exists between the theoretical concepts that physicists have formed about gases and other ponderable bodies, and Maxwell's theory of electromagnetic processes in so-called empty space«, and he went on to say that the former was based on a finite number of atoms and electrons, and the latter on continuous spatial functions. He noted that Maxwell's theory had proved itself superbly in describing purely optical phenomena, but one should keep in mind that these are time averages rather than instantaneous values, and the use of continuous functions leads to contradictions when applied to the phenomena of emission and transformation of light.

Although chemists usually take the proof for the existence of photons from that part of Einstein's paper which deals with the photoelectric effect (the emission of electrons from molecules or solids due to the absorption of electromagnetic radiation), Einstein's arguments for light quantisation are mainly based on the statistical analysis of the equilibrium between matter and electromagnetic radiation, and in particular the radiation which we refer to as black-body.

Black-body radiation was defined in 1862 by Kirchhoff to be the radiation in equilibrium with a body which absorbs and emits radiation uniformly at all frequencies. The energy of such radiation is independent of the nature of the emitting body, being only a function of the temperature and frequency. It took nearly forty years of experimentation to establish the precise form of the black-body density function. Its most important feature is that it has a single maximum, which moves to higher frequencies as the temperature is increased.

At the end of the 19th century the theory of black-body radiation was in a very unsatisfactory state. In 1893, Wien had derived a formula for the energy density which was in agreement with experiment for high frequencies, but not for low frequencies, and in 1900 Lord Rayleigh produced a formula (later amended in collaboration with Jeans) that was correct for low frequencies but not for high frequencies. The foundations for a correct theory were given by Planck in a paper presented to the German Physical Society in December 1900, a date which is usually taken as the birth of quantum mechanics.

Planck followed the statistical methods developed earlier by Boltzmann, and used his important formula for the entropy:

$$S = k \ln W \tag{4}$$

With a counting procedure for W that predates the Bose-Einstein statistics given 25 years later. He divided the total energy into a finite number of elements of equal magnitude, and counted the number of ways of distributing these energies amongst the individual oscillators. If the size of the energy elements is set equal to $h\nu$ (h later to be called Planck's constant), then an expression for the black-body radiation density was obtained that he had deduced earlier by interpolating the Wien and Rayleigh-Jeans laws. As Einstein pointed out in 1906, Planck's method of counting implies that the oscillator energies are restricted to the formula:

$$E_n = nh\nu \tag{5}$$

where n is an integer, and with this assumption an oscillator can only take up or lose energy in multiples of $h\nu$. Planck showed that the average energy of an oscillator is:

$$\varepsilon = h\nu/(\exp(h\nu/kT) - 1) \tag{6}$$

where k is Boltzmann's constant, and the energy-density of the radiation in equilibrium with this is:

$$P = (8\pi\, \nu^2/c^3)\, \varepsilon \tag{7}$$

When it was produced Planck's formula was considered as little more than a highly accurate representation of the data, and only later did its position in the foundations of quantum physics become obvious.

Einstein first demonstrated that classical statistical mechanics and Maxwell's electromagnetic theory lead irrevocably to the Rayleigh-Jeans law for black-body radiation. He then showed that if one adopted Wien's law one deduced that the entropy of the radiation depends on the volume in the same way as the entropy of an ideal gas depends on volume. This is rather slender evidence for reaching the conclusion that monochromatic radiation of low intensity behaves thermodynamically as if it contains independent energy quanta, and using Planck's formula for the radiation density these quanta must have an energy hν. Einstein then took a much bolder step by saying: »it seems reasonable to investigate whether the laws governing the emission and transformation of light are also constructed as if light consisted of such quanta«. In 1909 Einstein attended his first physics conference and referred to his light quanta as point-like objects. The word photon to describe these was coined by G. N. Lewis in 1926.

The photoelectric effect provides a much clearer proof of the existence of light quanta and of their energy. The subject was in its infancy when Einstein explained its most important features, and it needed a further ten years before his theory was confirmed. The first observations were by Hertz in 1887. He noted that sparks produced at a metal surface generated secondary sparks at a neighbouring surface, and he showed that the length of the secondary sparks was reduced if the light from the primary sparks was filtered out. He later showed that light alone could produce sparks. In 1899 J. J. Thomson showed that particles emitted from a metal by ultra-violet light were electrons (by measuring their e/m ratio), and in 1902 Lenard made the important discovery that the energy of the emitted electrons was independent of the light intensity. The only knowledge about the role of the light frequency at that time was that the electron energy increased as the frequency increased.

Einstein proposed that a light quantum could give all its energy to a single electron, and that some energy would be lost in travelling to the surface with a further loss when it was emitted. Electrons at the surface would come off with maximum energy, and the expression for this, in modern notation, would be

$$E = h\nu - P \tag{8}$$

where P is called the work function of the metal (or the ionisation potential in the case of electron emission from an individual molecule). Einstein showed that the energy balance in this equation agreed in order of magnitude with Lenard's results.

The fact that E varies linearly with ν, and that the slope is a universal constant, h, was not known by experiment in 1905, and it was not until a paper by Millikan in 1916 that this was confirmed, and a value determined for h from such experiments; 6.57×10^{-27} J s (modern value 6.626×10^{-27} J s).

The photoelectron effect showed that the energy of radiation was transmitted in quanta. Another important line of experiment was on the scattering of X-Rays from substances of low molecular weight, for which it had been noted that secondary X-Rays were produced which had longer wavelengths, and in 1922 Compton suggested that when an X-Ray was scattered it gave all of its energy and momentum to an electron, and this in turn radiated a new X-Ray with lower frequency He made this proposal to explain some observations by others on secondary X-Ray emission. Compton examined the scattering of X-Rays on light elements and noted that the increase in wavelength was greater at large angles of scattering than at small angles. As the energy of an X-Ray is so much larger than the binding energy of an electron in light elements, these observations could essentially be treated as the collision of an X-Ray with a free electron, and Compton showed that the scattering angle and wavelength shift could be effectively described by classical (relativistic) mechanics. The recoil electrons were later examined by Wilson using his cloud expansion technique, and these were also found to fit Compton's mechanics. Compton scattering showed that momentum as well as energy was transmitted in quanta, thus confirming Einstein's proposal that light was composed of individual particles, later to be called photons.

In a paper published in 1909 Einstein said: »It is my opinion that the next phase in the development of theoretical physics will bring us a theory of light that can be interpreted as a kind of fusion of the wave and emission (particle) theory«. This proposal for wave-particle duality led de Broglie later to postulate that a similar duality also applied to matter. Einstein's key role in both the quantisation of radiation and, through his interpretation of Planck's formula, of matter energies, probably places him as the leading originator of quantum theory; it is odd that he spent the rest of his life trying to explain why he was not happy with its formulation or its philosophical implications.

END PIECE

Science never ends but articles must. Quantum mechanics as formulated by Heisenberg and Schrödinger, and interpreted by many others, was combined with relativity by Dirac. Bringing electromagnetic radiation into the formalism led Tomanaga, Schwinger and Feynman to develop quantum electrodynamics (QED). It has been said that QED is the most highly tested of all physical theories, which reached its peak when Bethe and then Tomanaga showed that it explained the Lamb shift, this is the observation of Lamb and Rethorford in 1947 that the 2s level of the hydrogen atom has a slightly higher energy than the 2p; the two are degenerate in Schrödinger and Dirac mechanics. In QED the photon appears as the carrier of the electromagnetic interaction between charged particles; this interaction is interpreted as the exchange of photons. The mathematical structure of QED was later expanded to quantum chromodynamics which was needed to explain the strong interaction between quarks. But I am going well beyond by knowledge. It must be said however, that anyone who believes that we have come to the end of the story of light has ignored the pattern of scientific advances. The next steps will, as always, be fascinating.

REFERENCES

In writing this article I relied mainly on secondary texts, going back to some primaries only when there was disagreement about key facts (like whether Roemer deduced a value for the velocity of light). For this reason I only give a bibliography of books that provided most of my information, and have mentioned primary sources only in the text. The Internet was also very important in searching for my material, but although there is an enormous amount available, it rarely provides complete coverage of a topic, and is subject to the inaccuracies of any unrefereed communications.

1. E. Whittaker, *A History of the Theories of Aether and Electricity*, Thomas Nelson, Edinburgh, Vol. 1 (revised and enlarged), 1951; Vol. 2, 1953.
2. Newton's *Opticks* (with a preface by Bernard Cohen and an introduction by Edmund Whittaker), Dover Publications, New York, 1979.
3. A. Pais, *Subtle is the Lord*, *The Science and the Life of Albert Einstein*, Oxford UP, 1982.
4. J. Stachel (Ed.), *Einstein's Miraculous Year* (Five papers that changed the face of physics, Foreword by R. Penrose), Princeton UP, 1998.
5. S. F. Mason, *A History of the Sciences*, Collier Books, New York, 1962.
6. J. H. Saunders, *Velocity of Light*, Pergamon Press, Oxford, 1965.
7. K. D. Frome and L. Essen, *The Velocity of Light and Radio Waves*, Academic Press, New York, 1969.

SAŽETAK

Osvjetljavanje svjetla

John Murrell

Dan je povijesni pregled svojstava svjetla i razvoj koncepcija koje su omogućile razumijevanje tih svojstava.