# 1 History of ideas

Why should a textbook on physics begin with history? Why not start with what is known now and refrain from all the distractions of out-of-date material? These questions would be justifiable if physics were a complete and finished subject; only the final state would then matter and the process of arrival at this state would be irrelevant. But physics is not such a subject, and optics in particular is very much alive and constantly changing. It is important for the student to study the past as a guide to the future. Much insight into the great minds of the era of classical physics can be found in books by Magie (1935) and Segré (1984).

By studying the past we can sometimes gain some insight – however slight – into the minds and methods of the great physicists. No textbook can, of course, reconstruct completely the workings of these minds, but even to glimpse some of the difficulties that they overcame is worthwhile. What seemed great problems to them may seem trivial to us merely because we now have generations of experience to guide us; or, more likely, we have hidden them by cloaking them with words. For example, to the end of his life Newton found the idea of 'action at a distance' repugnant in spite of the great use that he made of it; we now accept it as natural, but have we come any nearer than Newton to understanding it? It is interesting that the question of 'action at a distance' resurfaced in a different way in 1935 with the concept of 'entangled photons', which will be mentioned in §1.7.2 and discussed further in §14.3.3.

The history of optics is summarized in Fig. 1.1, which shows many of the important discoveries and their interactions, most of which are discussed in the chapters that follow. First, there was the problem of understanding the nature of light; originally the question was whether light consisted of massive corpuscles obeying Newtonian mechanics, or was it a wave motion, and if so in what medium? As the wave nature became clearer, the question of the medium became more urgent, finally to be resolved by Maxwell's electromagnetic theory and Einstein's theory of relativity. But the quantum nature of physics re-aroused the wave–particle controversy in a new form, and today many basic questions are still being asked about the interplay between particle and wave representations of light.

We shall touch on some of these questions, which have been addressed by some very thought-provoking experiments, in Chapter 14.

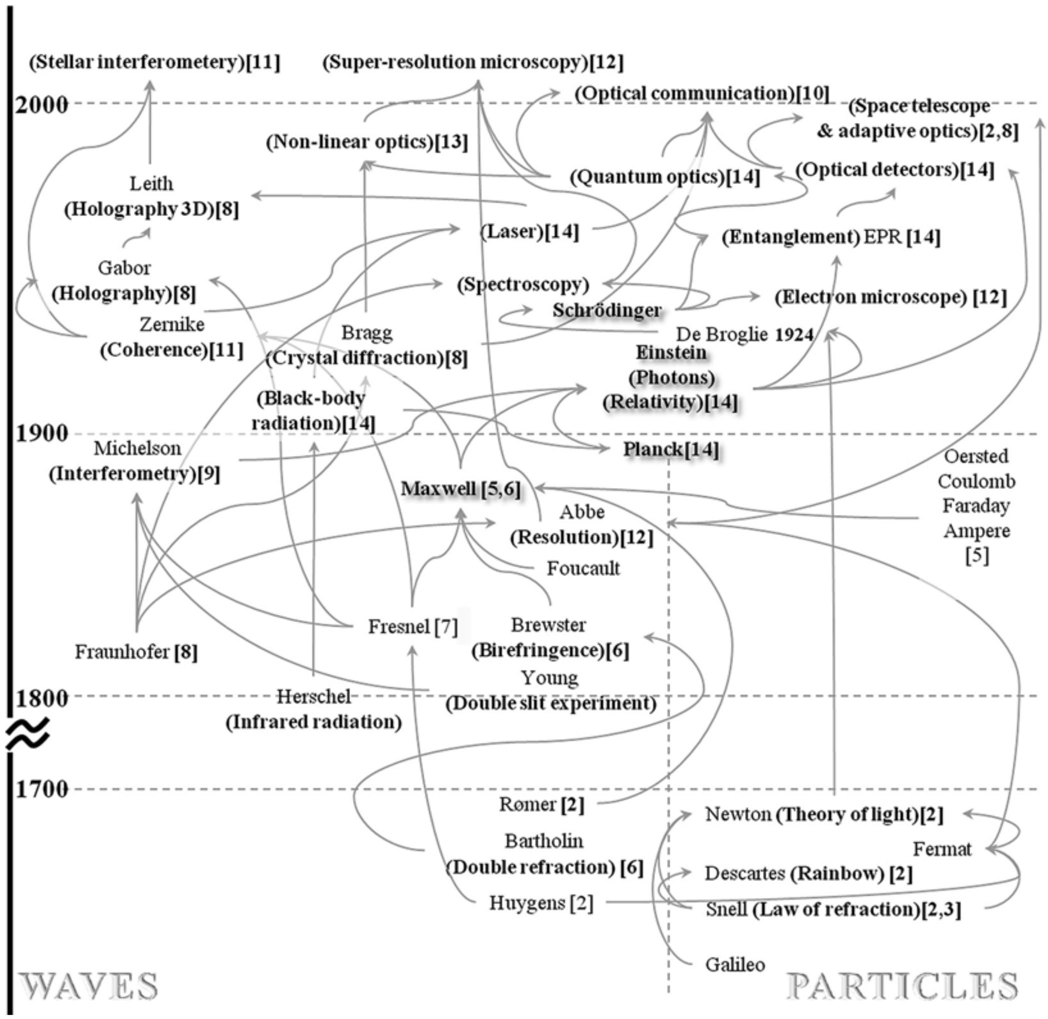


**Figure 1.1** The development of optics, showing many of the interactions. Notice that there was little development in the eighteenth century, mainly because of Newton's erroneous idea of light particles. The numbers in square brackets indicate the chapters where the topics are discussed.

A complementary trail follows the applications of optics. Starting with simple refractive imaging devices, well explained by corpuscular considerations, the wave theory became more and more relevant as the design of these instruments improved, and it became clear that bounds to their performance existed. But even the wave theory is not quite adequate to deal with the sensitivity of optical instruments, which is eventually limited by quantum theory. A fuller understanding of this is leading us today towards more sensitive and more accurate measurement and imaging techniques.

# 1.1 The nature of light

## 1.1.1 The basic facts

Let us go back to the time of Galileo (1564–1642). What was known about light in the seventeenth century? First of all, it travelled in straight lines and Galileo, who originated the idea of testing theories by experiment, tried unsuccessfully to measure its speed. Second, it was reflected off smooth surfaces and the laws of reflection were known. Third, it changed direction when it passed from one medium to another (refraction, §2.6.2); the laws for this phenomenon were not so obvious, but they were established by Snell (1591–1626) and were later confirmed by Descartes (1596–1650). Fourth, what we now call Fresnel diffraction (§7.2) had been discovered by Grimaldi (1618–63) and by Hooke (1635–1703). Finally, double refraction (§6.5) had been discovered by Bartholinus (1625–98). It was on the basis of these phenomena that a theory of light had to be constructed.

The last two facts were particularly puzzling. Why did shadows reach a limiting sharpness as the size of the source became small, and why did fringes appear on the light side of the shadow of a sharp edge? And why did light passing through a crystal of calcite (see Fig. 1.4) produce two images while light passing through most other transparent materials produced only one?

## 1.1.2 The wave–corpuscle controversy

Newton did consider gravitational bending of light. He obtained a value a factor of two smaller than later predicted by relativity, but this was not discovered till 1919 (§2.8)!

Two explanations were put forward: corpuscules and waves, and an acrimonious controversy resulted. Newton (1642–1727) threw his authority behind the theory that light is corpuscular, mainly because his first law of motion said that if no force acts on a particle it will travel in a straight line; he assumed that the velocity of the corpuscles was large enough that gravitational bending would be negligible. Double refraction he explained by some asymmetry in the corpuscles, so that their directions depended upon whether they passed through the crystal forwards or sideways. He envisaged the corpuscles as resembling magnets and the word 'polarization' is still used even though this explanation has long been discarded.

Diffraction, however, was difficult. Newton realized its importance and was aware of what are now known as Newton's rings (§9.1.2), and he saw that the fringes formed in red light were separated more than those formed in blue light. He was also puzzled by the fact that light was partly transmitted and partly reflected by a glass surface; how could his corpuscles sometimes go through and sometimes be reflected? He answered this question by propounding the

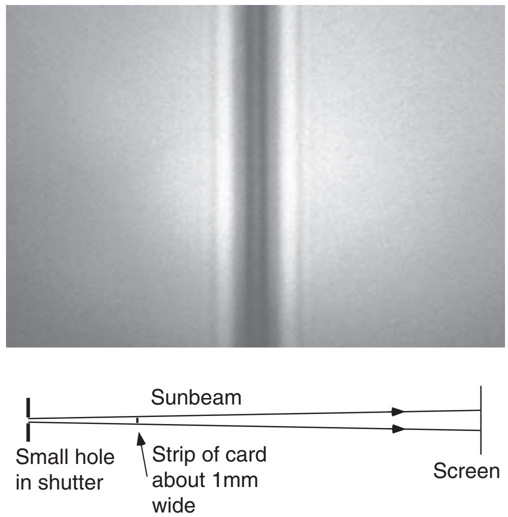


**Figure 1.2**
Young's interference experiment. In a narrow beam of sunlight he placed a narrow strip of card, about 1 mm in width, to create two separate beams, and then observed on a screen that there were fringes in the region where the two beams overlapped.

idea that they had internal vibrations that caused 'fits of reflexion' and 'fits of transmission'; in a train of corpuscles some would go one way and some the other. He even worked out the lengths of these 'fits' (which came close to what we now know as half the wavelength). But the idea was very cumbersome and was not really satisfying.

His contemporary Huygens (1629–95) was a supporter of the wave theory. With it he could account for diffraction and for the behaviour of two sets of waves in a crystal, without explaining *how* the two sets arose. Both he and Newton thought that light waves, if they existed, must be like sound waves, which are longitudinal. It is surprising that two of the greatest minds in science should have had this blind spot; if they had thought of transverse waves, the difficulties of explaining double refraction would have disappeared.

### 1.1.3 Triumph of the wave theory

Newton's authority kept the corpuscular theory going until the end of the eighteenth century, but by then ideas were coming forward that could not be suppressed. In 1801 Young (1773–1829) demonstrated interference fringes between waves from two sources (Fig. 1.2) – an experiment so simple to carry out and interpret that the results were incontrovertible. In 1815 Fresnel (1788–1827) worked out the theory of the Grimaldi–Hooke fringes (§7.1) and in 1821 Fraunhofer (1787–1826) invented the diffraction grating and produced diffraction patterns in parallel light for which the theory was much simpler (§9.2). These three men laid the foundation of the wave theory that is still the basis of what is now called physical optics.

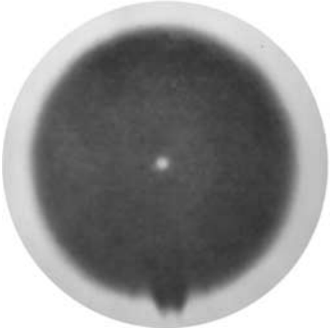

**Figure 1.3**
Fresnel and Arago's experiment: the bright spot at the centre of the shadow of a disc. The experimental arrangement was similar to that of Young, shown in Fig. 1.2.

The defeat of the corpuscular theory, at least until the days of quantum ideas, came in 1818. In that year, Fresnel wrote a prize essay on the diffraction of light for the French Académie des Sciences on the basis of which Poisson (1781–1840), one of the judges, produced an argument that seemed to invalidate the wave theory by *reductio ad absurdum*. Suppose that a shadow of a perfectly round object is cast by a point source; at the periphery all the waves will be in phase, and therefore the waves should also be in phase at the centre of the shadow, and there should therefore be a bright spot at this point. Absurd! Then Fresnel and Arago (1786–1853) carried out the experiment and found that there really was a bright spot at the centre (Fig. 1.3). The triumph of the wave theory seemed complete.

The Fresnel–Arago experiment is discussed in detail in §7.2.4.

## 1.2 Speed of light

The methods that Galileo employed to measure the speed of light were far too crude to be successful. In 1678 Römer (1644–1710) realized that an anomaly in the times of successive eclipses of the moons of Jupiter could be accounted for by a finite speed of light, and deduced that it must be about $3 \times 10^8$ m s$^{-1}$. In 1726 Bradley (1693–1762) made the same deduction from observations of the small ellipses that the stars describe in the heavens; since these ellipses have a period of one year they must be associated with the movement of the Earth.

It was not, however, until 1850 that direct measurements were made, by Fizeau (1819–96) and Foucault (1819–68), confirming the estimates obtained by Römer and Bradley. Knowledge of the exact value was an important confirmation of Maxwell's (1831–79) theory of electromagnetic waves (§5.1), which allowed the wave velocity to be calculated from the results of laboratory experiments on static and current electricity. In the hands of Michelson (1852–1931) their methods achieved a high degree of accuracy – about 0.03 per cent. Subsequently much more accurate determinations have been made, and the velocity of light in vacuum has now become one of the fundamental constants of physics, replacing the standard metre.

### 1.2.1 Refractive index

The idea that refraction occurs because the velocity of light is dependent on the medium dates back to Huygens and Newton. According to the corpuscular theory, the speed of light should be greater in a denser medium than in air because the corpuscles must be attracted towards the denser medium to account for the changed direction of the refracted light. According to the wave theory, the waves must travel more slowly in the medium and 'slew' round to give the new direction (Fig. 2.9). Foucault's method of measurement only required a relatively short path, and the speed of light could therefore be measured directly in media other than air – water, for example. Although the wave theory was by then completely accepted, Foucault provided welcome confirmation that the velocity was indeed smaller in water. A variation on the experiment performed by Fizeau provided a method of investigating the effects of motion of the medium on the velocity of light, because it was possible to carry out the measurements when the water was flowing through the apparatus (§9.4.1). The results could not be explained on the basis of nineteenth century physics of course, but preempted the theory of relativity.

## 1.3 The nature of light waves: Transverse or longitudinal?

The distinction between transverse and longitudinal waves had been appreciated early in the history of physics; sound waves were found to be longitudinal and water waves were obviously transverse. In the case of light waves, the phenomenon that enabled a decision to be made was that of double refraction in calcite. As we mentioned before, Huygens had pointed out that this property, which is illustrated in Fig. 1.4, means that the orientation of the crystal must somehow be related to some direction in the wave, but he had failed to appreciate the connection with transversality of the waves.

The greatest step towards understanding the waves came from a completely different direction – the theoretical study of magnetism and electricity.

In the first half of the nineteenth century the relationship between magnetism and electricity had been worked out fairly thoroughly, by men such as Oersted (1777–1851), Ampère (1775–1836) and Faraday (1791–1867). In order to visualize his experimental results, Faraday invented around 1851 the concept of 'lines of force', which described the 'action at a distance' that had so worried his predecessors in magnetism, electricity and gravitation. In 1865, Maxwell was inspired to combine his predecessors' observations in mathematical form by describing the region of influence around electric charges and magnets as an

The concept of a 'field', which is widely used today in all areas of physics, was originated by Faraday in this work.


**Figure 1.4**
**Double refraction in a calcite crystal.**

'electromagnetic field' and expressing the observations in terms of differential equations. In manipulating these equations he found that they could assume the form of a transverse wave equation (§2.1.1), a result that had already been guessed by Faraday in 1846. The velocity of the wave could be derived from the known magnetic and electric constants, and was found to be equal to the measured velocity of light; thus light was established as an electromagnetic disturbance. A key to Maxwell's success was his invention of the concept of a 'field', which is a continuous function of space and time representing the mutual influence of one body on another, a prolific idea that has dominated the progress of physics ever since then. This began one of the most brilliant episodes in physics, during which different fields and ideas were brought together and related to one another.

## 1.4 Quantum theory

With the marriage of geometrical optics and wave theory (physical optics) it seemed, up to the end of the nineteenth century, that no further rules about the behaviour of light were necessary. Nevertheless there remained some basic problems, as the study of the light emitted by hot bodies indicated. Why do such bodies become red-hot at about 600 °C and become whiter as the temperature increases? The great physicists such as Kelvin (1824–1907) were well aware of this problem, but it was not until 1900 that Planck (1858–1947) put forward, very tentatively, an ad hoc solution, now known as the quantum theory.

Planck had a hard time defending his doctoral dissertation, in which the idea of quantization was proposed!

Planck's idea (§14.1.1) was that wave energy is divided into packets (quanta), now called photons, whose energy content is proportional to the frequency. The lower frequencies, such as those of red light, are then more easily produced than higher frequencies. The idea was not liked – even Planck himself was hesitant in proposing it – but gradually scepticism was overcome as more and more experimental evidence in its favour was produced. By about 1920 it was generally accepted, largely on the basis of Einstein's (1879–1955) study of the photo-electric effect (1905) and of Compton's (1892–1962) understanding of energy and momentum conservation in the scattering of X-rays by electrons (1923); even though, in retrospect, neither of these experiments conclusively shows that an electromagnetic wave itself is quantized, but only that it interacts with a material in a quantized way, which might be a property of the material itself. The real proof had to wait for the advent of non-linear optics (§1.7.2).

### 1.4.1 Wave–particle duality

So it seems that light has both corpuscular properties and wave-like features at the same time. This duality is still difficult to appreciate to those of us who like intuitive physical pictures. The energy of a wave is distributed through space; the energy of a particle would seem to be concentrated in space. A way of understanding duality questions in linear optics is to appreciate that the wave intensity tells us the probability of finding a photon at any given point. The corpuscular features only arise when the wave interacts with a medium, such as a detector, and gives up its energy to it. Thus, any given problem should be solved in terms of wave theory right until the bitter end, where the outcome is detected. However, this interpretation is not sufficient when non-linear phenomena are involved; curious correlations between different photons then arise, defying attempts to make simple interpretations (§14.3).

### 1.4.2 Corpuscular waves

As usual in physics one idea leads to another, and in 1924 a new idea occurred to de Broglie (1892–1987), based upon the principle of symmetry. Faraday had used this principle in his discovery of electromagnetism; if electricity produces magnetism, does magnetism produce electricity? De Broglie asked, 'If waves are corpuscles, are corpuscles waves?' Within three years his question had been answered. Davisson (1881–1958) and Germer (1896–1971) by ionization methods and G. P. Thomson (1892–1975) by photographic methods, showed that fast-moving electrons could be diffracted by matter similarly to X-rays. Since then other particles such as neutrons, protons and atoms have also been diffracted. Based on these experiments, Schrödinger (1887–1961) in 1928 produced a general wave theory of matter, which has stood the test of time down to atomic dimensions at least.

## 1.5 Optical instruments

### 1.5.1 The telescope

Newton apparently did not realize that different types of glass had different degrees of dispersion, so he did not think that an achromatic doublet could be made.

Although single lenses had been known from time immemorial, it was not until the beginning of the seventeenth century that optical instruments as we know them came into being. Lippershey (d. 1619) discovered in 1608, probably accidentally, that two separated lenses, an objective and an eye lens, could produce a clear enlarged image of a distant object (§3.3.2). Galileo seized upon the discovery, made his own telescope, and began to make a series of discoveries – such as Jupiter's moons and Saturn's rings – that completely altered the subject of astronomy. Newton, dissatisfied with the colour defects in the image, invented the reflecting telescope (Fig. 1.5).

Modern telescopes have retained the basic elements of these original designs, but many additional features have made them much more powerful and accurate. In about 1900, Lord Rayleigh (1842–1919) showed that the angular resolution of a telescope is limited by diffraction at its aperture (§12.2.1), so that bigger and bigger telescopes were built in order to produce brighter images and, hopefully, to improve the resolution too. But it appeared that resolution was limited by atmospheric turbulence effects once the aperture diameter exceeded about 15 cm. Both Rayleigh's resolution limit and the atmospheric limitation were first circumvented by Michelson in 1921, who used interference between a pair of small telescope apertures (15 cm diameter) separated by several metres, to achieve resolution equivalent to the separation, and not the telescope aperture (§11.8.1). Later, in 1972, Labeyrie showed how to overcome the atmospheric limitations of a single large-aperture telescope, and for the first time achieved diffraction-limited resolution from the Palomar 2.5 m ground-based telescope by using an image-combination technique called 'speckle interferometry' (§12.7).

The story of how the Hubble telescope was launched with a serious aberration in the primary mirror, and how this was analyzed and corrected *in situ*, is told in §8.9.

Since 1994, superb astronomical images with almost diffraction-limited resolution are being routinely obtained with the Hubble Space Telescope, which has an aperture of 2.4 m and is of course not limited by atmospheric turbulence or transmission. But more recently, ground-based telescopes with apertures up to 10 m diameter use real-time atmospheric correction at infra-red and visible wavelengths, called 'adaptive optics', to produce stellar images that rival those from the space telescope in brightness and resolution.

### 1.5.2 The microscope

The story of the microscope is different. Its origin is uncertain; many people contributed to its early development. The microscope originated from the

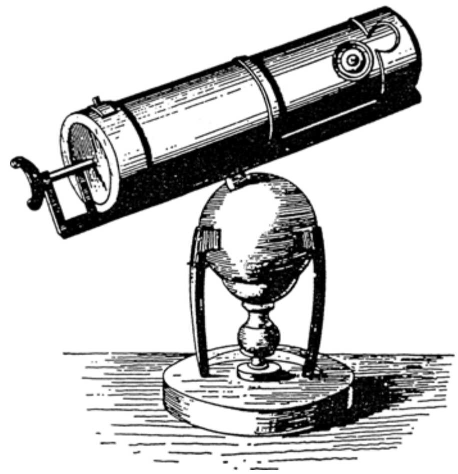

**Figure 1.5**

**Newton's reflecting telescope.**

**Figure 1.6**

**Hooke's microscope, from his *Micrographia*.**

magnifying glass. In the sixteenth and seventeenth centuries considerable ingenuity was exercised in making high-powered lenses; a drop of water or honey could produce wonderful results in the hands of an enthusiast. Hooke (1635–1703) played perhaps the greatest part in developing the compound microscope which consisted, like the telescope, of an objective and an eye lens (§3.4). Some of his instruments (Fig. 1.6) already showed signs of future trends in design. One can imagine the delight of such an able experimenter in having the privilege of developing a new instrument and of using it to examine for the first time the world of the very small, depicted in his *Micrographia* (1665). Microscope technology improved continuously throughout the years, producing ever clearer images of absorbing objects, but an invention by Zernike (1888–1966)

We suggest you try making your own 'Hooke microscope' using a drop of honey or better, corn syrup, and relive some of Hooke's discoveries.