1 Light as Waves, Rays and Photons

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

Isaac Newton, *Opticks*

All these 50 years of conscious brooding have brought me no nearer to the answer to the question 'What are light quanta?'. Nowadays every Tom, Dick and Harry thinks he knows it, but he is mistaken.

Albert Einstein, *A Centenary Volume,* 1951.

How wonderful that we have met with a paradox. Now we have some chance of making progress.

Niels Bohr (quoted by L.I. Ponomarev in *The Quantum Dice*).

Light is an *electromagnetic wave*: light is emitted and absorbed as a stream of discrete *photons*, carrying packets of energy and momentum. How can these two statements be reconciled? Similarly, while light is a wave, it nevertheless travels along straight lines or rays, allowing us to analyse lenses and mirrors in terms of geometric optics. Can we use these descriptions of waves, rays and photons interchangeably, and how should we choose between them? These problems, and their solutions, recur throughout this book, and it is useful to start by recalling how they have been approached as the theory of light has evolved over the last three centuries.

1.1 The Nature of Light

In his famous book *Opticks*, published in 1704, Isaac Newton described light as a stream of particles or corpuscles. This satisfactorily explained rectilinear propagation, and allowed him to develop theories of reflection and refraction, including his experimental demonstration of the splitting of sunlight into a spectrum of colours by using a prism. The particles in rays of different colours were supposed to have different qualities, possibly of mass, or size or velocity. White light was made up of a compound of coloured rays, and the colours of transparent materials were due to selective absorption. It was, however, more difficult for him to explain the coloured interference patterns in thin films, which we now call Newton's rings (see Chapter 9). For this, and for the partial reflection of light at a glass surface, he suggested a kind of periodic motion induced by his corpuscles, which reacted on the particles to give 'fits of easy reflection and transmission'. Newton also realized that double refraction in a calcite crystal (Iceland spar) was best explained by attributing a rectangular

Optics and Photonics: An Introduction, Second Edition F. Graham Smith, Terry A. King and Dan Wilkins

cross-section (or 'sides') to light rays, which we would now describe as *polarization* (Chapter 7). He nevertheless argued vehemently against an actual wave theory, on the grounds that waves would spread in angle rather than travel as rays, and that there was no medium to carry light waves from distant celestial bodies.

The idea that light was propagated as some sort of wave was published by René Descartes in *La Dioptrique* (1637); he thought of it as a pressure wave in an elastic medium. Christiaan Huygens, a Dutch contemporary of Newton, developed the wave theory; his explanation of rectilinear propagation is now known as 'Huygens' construction'. He correctly explained refraction in terms of a lower velocity in a denser medium. Huygens' construction is still a useful concept, and we use it later in this chapter.

It was not, however, until 100 years after Newton's *Opticks* that the wave theory was firmly established and the wavelength of light was found to be small enough to explain rectilinear propagation. In Thomas Young's double slit experiment (see Chapter 8), monochromatic light from a small source passed through two separate slits in an opaque screen, creating interference fringes where the two beams overlapped; this effect could only be explained in terms of waves. Augustin Fresnel, in 1821, then showed that the wave must be a *transverse* oscillation, as contrasted with the longitudinal oscillation of a sound wave; following Newton's ideas of rays with 'sides', this was required by the observed polarization of light as in double refraction. Fresnel also developed the theories of partial reflection and transmission (Chapter 5), and of diffraction at shadow edges (Chapter 10). The final vindication of the wave theory came with James Clerk Maxwell, who synthesized the basic physics of electricity and magnetism into the four Maxwell equations, and deduced that an electromagnetic wave would propagate at a speed which equalled that of light.

The end of the nineteenth century therefore saw the wave theory on an apparently unassailable foundation. Difficulties only remained with understanding the interaction of light with matter, and in particular the 'blackbody spectrum' of thermal radiation. This was, however, the point at which the corpuscular theory came back to life. In 1900 Max Planck showed that the form of the blackbody spectrum could be explained by postulating that the walls of the body containing the radiation consisted of harmonic oscillators with a range of frequencies, and that the energies of those with frequency ν were restricted to integral multiples of the quantity $h\nu$. Each oscillator therefore had a fundamental energy quantum

$$\boxed{E = h\nu} \tag{1.1}$$

where h became known as *Planck's constant*. In 1905 Albert Einstein explained the photoelectric effect by postulating that electromagnetic radiation was itself quantized, so that electrons are emitted from a metal surface when radiation is absorbed in discrete quanta. It seemed that Newton was right after all! Light was again to be understood as a stream of particles, later to become known as photons. What had actually been shown, however, was that light energy and the momentum carried by a light wave existed in discrete units, or quanta; photons should be thought of as events at which these quanta are emitted or absorbed.

If light is a wave that has properties usually associated with particles, could material particles correspondingly have wave-like properties? This was proposed by Louis de Broglie in 1924, and confirmed experimentally three years later in two classical experiments by George Thomson and by Clinton Davisson and Lester Germer. Both showed that a beam of particles, like a light ray encountering an obstacle, could be diffracted, behaving as a wave rather than a geometric ray. The diffraction pattern formed by the spreading of an electron beam passing through a hole in a metal sheet, for example, was the same as the diffraction pattern in light which we explore in Chapter 10. Furthermore, the wavelength λ involved was simply related to the momentum p of the electrons by

$$\boxed{\lambda = \frac{h}{p}.} \tag{1.2}$$

The constant h was again Planck's constant, as in the theory of quanta in electromagnetic radiation; for material waves λ is the *de Broglie* wavelength. A general wave theory of the behaviour of matter, *wave mechanics*, was developed in 1926 by Erwin Schrödinger following de Broglie's ideas. Wave mechanics revolutionized our understanding of how microscopic particles were described and placed limitations on the extent of information one could have about such systems – the famous Heisenberg uncertainty relationship.

The behaviour of both matter and light evidently has dual aspects: they are in some sense both particles and waves. Which aspect best describes their behaviour depends on the circumstances; light propagates, diffracts and interferes as a wave, but is emitted and absorbed discontinuously as photons, which are discrete packets of energy and momentum. Photons do not have a continuous existence, as does for example an electron in the beam of an accelerator machine; in contrast with a material particle it is not possible to say where an individual photon is located within a light beam. In some contexts we nevertheless think of the light within some experimental apparatus, such as a cavity or a laser, as consisting of photons, and we must then beware of following Newton and being misled by thinking of photons as particles with properties like those of material particles.

Although photons and electrons have very similar wave-like characteristics, there are several fundamental differences in their behaviour. Photons have zero mass; the momentum p of a photon in equation (1.1) is related to its kinetic energy E by $E = pc$, as compared with $E = p^2/2m$ for particles moving well below light speed. Unlike electrons, photons are not conserved and can be created or destroyed in encounters with material particles. Again, their statistical behaviour is different in situations where many photons or electrons can interact, as for example the photons in a laser or electrons in a metal. No two electrons in such a system can be in exactly the same state, while there is no such restriction for photons: this is the difference between Fermi–Dirac and Bose–Einstein statistics respectively for electrons and for photons.

In the first two-thirds of this book we shall be able to treat light mainly as a wave phenomenon, returning to the concept of photons when we consider the absorption and emission of electromagnetic waves.

1.2 Waves and Rays

We now return to the question: how can light be represented by a ray? Huygens' solution was to postulate that light is propagated as a wavefront, and that at any instant every point on the wavefront is the source of a wavelet, a secondary wave which propagates outward as a spherical wave (Figure 1.1)

Each wavelet has infinitesimal amplitude, but on the common envelope where countless wavelets intersect, they reinforce each other to form a new wavefront of finite amplitude. In this way, successive positions of the wavefront can be found by a step-by-step process. The envelope[1] of the

[1]To define the envelope evolved after a short time from a wavefront segment, take a finite number N of wavelets with evenly spaced centres, and note the intersection points between adjacent wavelets. In the limit that N goes to infinity, the intersection points crowd together and constitute the envelope, which is the new wavefront.

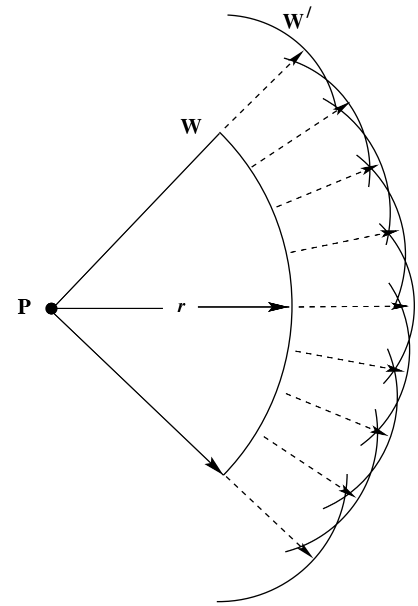

Figure 1.1 Huygens' secondary wavelets. A spherical wavefront W has originated at P and after a time t has a radius $R = ct$, where c is the speed of light. Huygens' secondary wavelets originating on W at time t combine to form a new wavefront W' at time t', when the radii of the wavelets are $c(t' - t)$

wavelets is perpendicular to the radius of each wavelet, so that the ray is the normal to a wavefront. This simple Huygens wavefront concept allows us to understand both the rectilinear propagation of light along ray paths and the basic geometric laws of reflection and refraction. There are obvious limitations: for example, what happens at the edge of a portion of the wavefront, as in Figure 1.1, and why is there no wave reradiated backwards? We return to these questions when we consider diffraction theory in Chapter 10.

Reflection of a plane wavefront W_1 reaching a totally reflecting surface is understood according to Huygens in terms of secondary wavelets set up successively along the surface as the wavefront reaches it (Figure 1.2(a)). These secondary wavelets propagate outwards and combine to form the reflected wavefront W_2. The rays are normal to the incident and reflected wavefronts. Light has travelled along each ray from W_1 to W_2 in the same time, so all path lengths from W_1 to W_2 via the mirror must be equal. The basic law of reflection follows: the incident and reflected rays lie in the same plane and the angles of incidence (i) and reflection (r) are equal.

Figure 1.2(b) shows the same reflection in terms of rays. Here we may find the same law of reflection as an example of *Fermat's principle of least time*, which states that the *time* of propagation is a minimum (or more strictly either a maximum or a minimum) along a ray path.[2] It is easy to see that the path of a light ray between the two points A and B (Figure 1.2 (b)) is a minimum if the angles i, r are equal. The proof is simple: construct the mirror image A′ of A in the reflecting surface, when the line A′PB must be straight for a minimum distance. Any other path AP′B is longer.

[2]This explanation of the basic law of reflection was first given by Hero of Alexandria (First century AD).

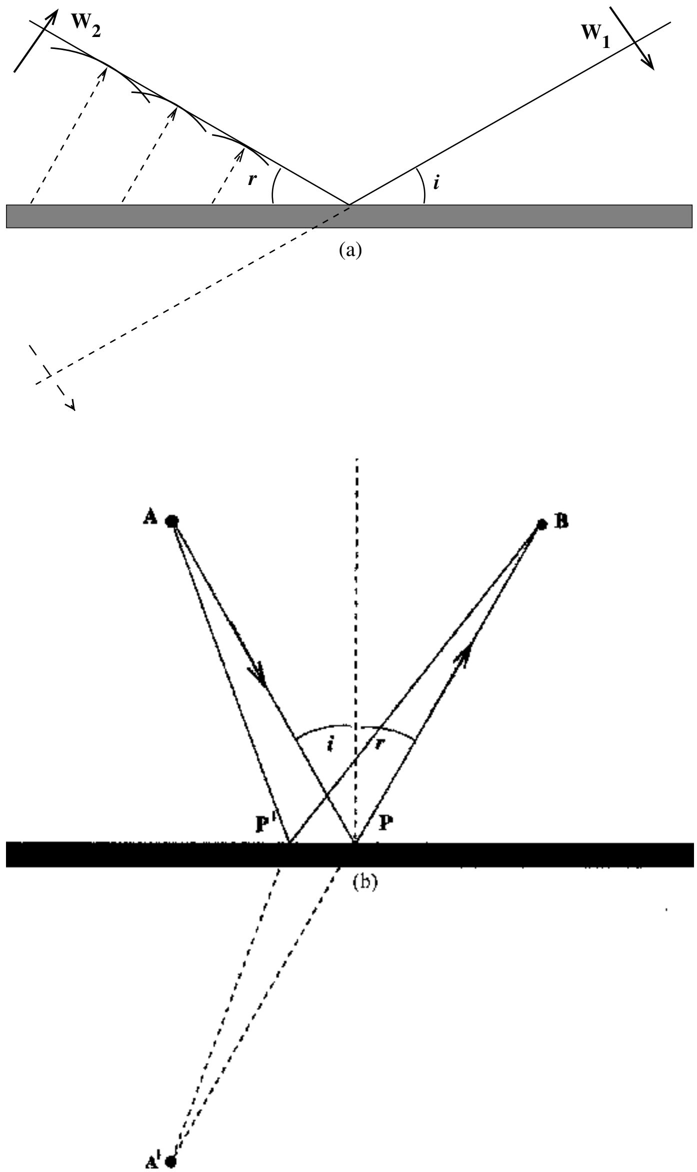

Figure 1.2 Reflection at a plane surface. (a) Huygens wave construction. The reflected wave W_2 is made up of wavelets generated as successive points on the incident plane wave W_1 reach the surface. (b) Fermat's principle. The law of reflection is found by making the path of a reflected light ray between the points A and B a minimum

Why are these two approaches essentially the same? Fermat tells us that the time of travel is the same along all paths close to an actual ray. In terms of waves this means that waves along these paths all arrive together, and reinforce one another as in Huygens' construction. When we consider periodic waves, we will express this by saying that they are *in phase*.

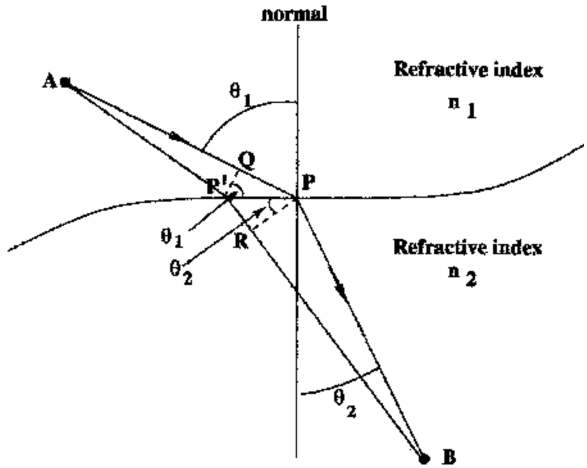

Figure 1.3 Refraction at a surface between transparent media with refractive indices n_1 and n_2. We assume the light rays and the surface normal all lie in the plane of the paper. Snell's law corresponds to a stationary value of the optical path $n_1\mathrm{AP} + n_2\mathrm{PB}$ between the fixed endpoints A, B; for small virtual variations such as shifting the point P to P′, the optical path changes negligibly

The basic law of refraction (Snell's law) may be found by applying either Huygens' or Fermat's principles to a boundary between two media in which the velocities of propagation v_1, v_2 are different; as Huygens realized, his secondary waves must travel more slowly in an optically denser medium. The *refractive indices* are defined as $n_1 = c/v_1, n_2 = c/v_2$ where c is the velocity of light in free space. As we now show, the Fermat approach shown in Figure 1.3 leads to Snell's law via some simple trigonometry.

The Fermat condition is that the travel time $(n_1\mathrm{AP} + n_2\mathrm{PB})/c$ is stationary (minimum, maximum, or point of inflection); this means that for any small change in the light path of order ϵ, the change in travel time vanishes as ϵ^2 (or even faster). The distance $n_1\mathrm{AP}+n_2\mathrm{PB}$ is called the optical path. We consider a small virtual displacement of the light rays from APB to AP′B. Denote the length PP′ as ϵ. By dropping perpendiculars from P and P′, we create two thin triangles AP′Q and BPR that become perfect isosceles triangles in the limit of zero displacement. Fermat requires then that the change of the optical path satisfies[3]

$$n_1\mathrm{QP} - n_2\mathrm{P'R} = n_1\epsilon\sin\theta_1 - n_2\epsilon\sin\theta_2 = \mathrm{O}(\epsilon^2). \tag{1.3}$$

Dividing by ϵ, and going to the limit $\epsilon = 0$, this leads directly to Snell's law of refraction:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{1.4}$$

Notice that this derivation works for a smoothly curving surface of any shape.

In Chapter 5 we show how the laws of reflection and refraction may be derived from electromagnetic wave theory.

[3]The notation $\mathrm{O}(\epsilon^2)$ designates a quantity that varies as ϵ^2 in the limit of vanishing epsilon.

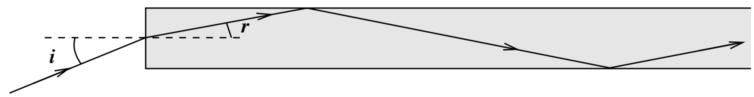

Figure 1.4 The light pipe. Rays entering at one end are totally internally reflected, and can be conducted along long paths which may include gentle curves

1.3 Total Internal Reflection

Referring again to Figure 1.3, and noting that the geometry is the same if the ray direction is reversed, we consider what happens if a ray inside the refracting medium meets the surface at a large angle of incidence θ_2, so that $\sin\theta_2$ is greater than n_1/n_2 and equation (1.4) would give $\sin\theta_1 > 1$. There can then be no ray above the surface, and there is *total internal reflection*. The internally reflected ray is at the same angle of incidence to the normal as the incident ray.

The phenomenon of total internal reflection is put to good use in the light pipe (Figure 1.4), in which light entering the end of a glass cylinder is reflected repeatedly and eventually emerges at the far end. The same principle is applicable to the transmission of light down thin optical fibres, but here the relation of the wavelength of light to the fibre diameter must be taken into account (Chapter 6).

1.4 The Light Wave

We now consider in more detail the description of the light wave, starting with a simple expression for a plane wave of any quantity ψ, travelling in the positive direction z with velocity v:

$$\boxed{\psi = f(z - vt).} \tag{1.5}$$

The function $f(z)$ describes the shape of ψ at the moment $t = 0$, and the equation states that the shape of ψ is unchanged at any later time t, with only a movement of the origin by a distance vt along the z axis (Figure 1.5). The minus sign in $(z - vt)$ indicates motion in the $+z$ direction; a plus sign would correspond to motion in the $-z$ direction. The variable quantity ψ may be a scalar, e.g. the pressure in a sound wave, or it may be a vector. If it is a vector, it may be *transverse*, i.e.

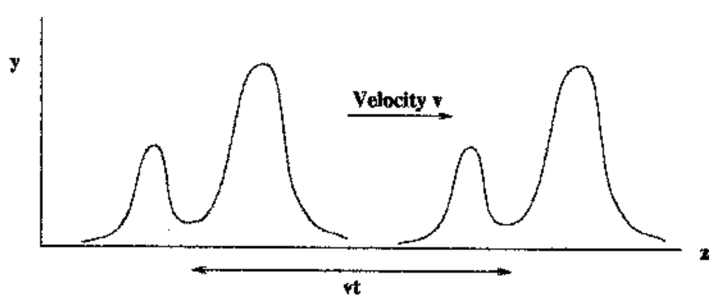

Figure 1.5 A wave travelling in the z direction with unchanging shape and with velocity v. At time $t = 0$ the waveform is $\psi = f(z)$, and at time t it is $\psi = f(z - vt)$

perpendicular to the direction of propagation, as are the waves in a stretched string, or the electric and magnetic fields in the electromagnetic waves which are our main concern. (These are the 'sides' which Newton attributed to his rays.) For most of optics it is sufficient to consider only the transverse electric field; indeed, as we shall see later, the results of scalar wave theory are sufficiently general that for many purposes we may just think of the magnitude of the electric field and forget about its vector nature.

At any one time the variation of ψ with z, i.e. the slope of the graph in Figure, 1.5, is $\partial\psi/\partial z$, and at any one place the rate of change of ψ is $\partial\psi/\partial t$. Changing to the variable $z' = (z - vt)$ and using the chain rule for partial differentiation:

$$\frac{\partial\psi}{\partial z} = \frac{\partial\psi}{\partial z'}\frac{\partial z'}{\partial z} = \frac{\partial\psi}{\partial z'} \tag{1.6}$$

$$\frac{\partial\psi}{\partial t} = \frac{\partial\psi}{\partial z'}\frac{\partial z'}{\partial t} = -v\frac{\partial\psi}{\partial z'}. \tag{1.7}$$

Similarly, the second differential of ψ with respect to z, i.e. $\partial^2\psi/\partial z^2$, which is the curvature of the graph in Figure 1.5, is related to the second differential with respect to time, i.e. the acceleration of ψ, by

$$\frac{\partial^2\psi}{\partial t^2} = v^2\frac{\partial^2\psi}{\partial z^2}. \tag{1.8}$$

This so-called one-dimensional wave equation applies to any wave propagating in the z direction with uniform velocity and without change of form.

The wave equation (1.8) may be extended to three dimensions, giving

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} \tag{1.9}$$

or in a more general and concise notation[4]

$$\boxed{\nabla^2\psi = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}.} \tag{1.10}$$

The form of the wave $f(z - vt)$ may be any continuous function, but it is convenient to analyse such behaviour in terms of *harmonic* waves, taking the simple form of a sine or cosine. (In Chapter 4 we show that any continuous function can be synthesized from the superposition of harmonic waves.) At any point such a wave varies sinusoidally with time t, and at any time the wave varies sinusoidally with distance z. The waveform is seen in Figure 1.6, which introduces the *wavelength* λ and *period* τ. At any point there is an oscillation with *amplitude* A. Equation (1.5) then becomes

$$\psi = A\sin\left[2\pi\left(\frac{z}{\lambda} - \frac{t}{\tau}\right)\right], \tag{1.11}$$

[4]Recall that ∇^2 is the *Laplacian operator:* $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$.

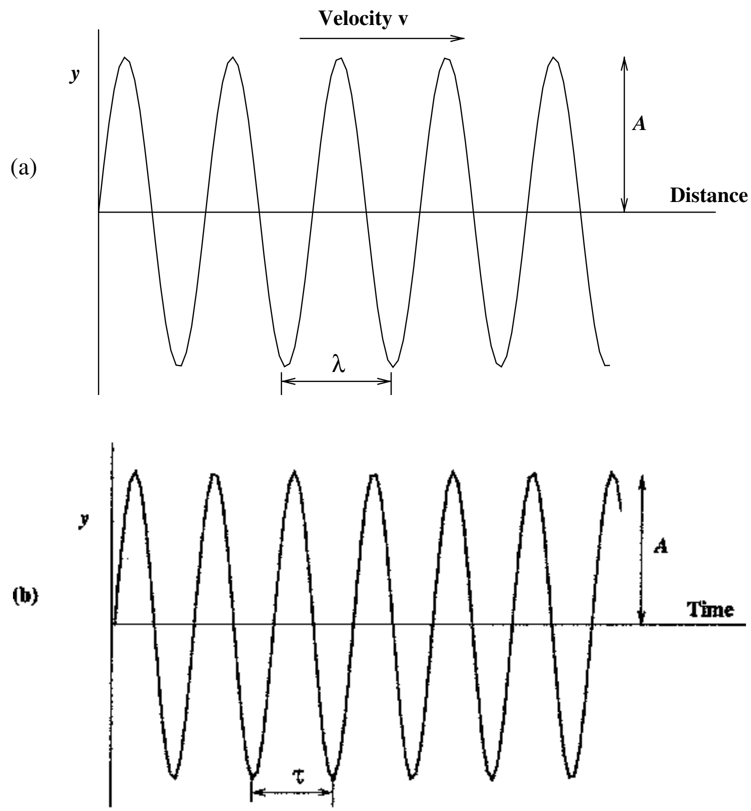

Figure 1.6 A progressive sine wave: (a) the wave at a fixed time; (b) the oscillation at a fixed point P

which is easily demonstrated to be a solution of the general wave equation (1.8) provided $\lambda/\tau = v$. The *frequency* of oscillation is $v = 1/\tau$. It is often convenient to use an *angular frequency* $\omega = 2\pi v$, and a *propagation constant* or *wave number*[5] $k = 2\pi/\lambda$. Equation (1.11) may then be written in terms of k as

$$\psi = A\sin(kz - \omega t). \tag{1.12}$$

The vector quantity $\mathbf{k} = (2\pi/\lambda)\hat{\mathbf{k}}$, where $\hat{\mathbf{k}}$ is the unit vector in the direction of $\mathbf{k}$, is also termed the *wave vector.*

Another powerful way of writing harmonic plane wave solutions of Equation (1.10) is in terms of complex exponentials

$$\psi = A\exp[\mathrm{i}(kz - \omega t)]. \tag{1.13}$$

Due to several elegant mathematical properties, including ease of differentiation and of visualization, complex functions like this can vastly simplify the process of combining waves of different amplitudes and phases, as we shall see in Chapter 4.

[5]Beware: the term *wave number* is also used in spectroscopy for $1/\lambda$, without the factor 2π.

1.5 Electromagnetic Waves

Although the idea that light was propagated as a combination of electric and magnetic fields was developed qualitatively by Michael Faraday, it required a mathematical formulation by Maxwell before the process could be clearly understood. In Chapter 5 we derive the electromagnetic wave equation from Maxwell's equations, and show that all electromagnetic waves travel with the same velocity in free space. There are two variables in an electromagnetic wave, the electric and magnetic fields **E** and **B**; both are vector quantities, but each can be represented by the variable ψ in the wave equation (1.10). As shown in Chapter 5, they are both transverse to the direction of propagation, and mutually perpendicular. Their magnitudes[6] are related by

$$\boxed{E = vB} \tag{1.14}$$

where v is the velocity of light in the medium. Since the electric and magnetic fields are mutually perpendicular and their magnitudes are in a fixed ratio, only one need be specified, and the magnitude and direction of the other follow. Equation (1.14) is true in general, but note that the velocity v in a dielectric such as glass is less than the free space velocity c; the *refractive index* n of the medium is

$$\boxed{n = \frac{c}{v}.} \tag{1.15}$$

As Huygens realized, light travels more slowly in dense media than in a vacuum.

In a transverse wave moving along a direction z the variable quantity is a vector which may be in any direction in the orthogonal plane x, y. The relevant variable for electromagnetic waves is conventionally chosen as the electric field **E**. The *polarization* of the wave is the description of the behaviour of the vector **E** in the plane x, y. The plane of polarization is defined as the plane containing the electric field vector and the ray, i.e. the z axis. If the vector **E** remains in a fixed direction, the wave is *linearly* or *plane* polarized; if the direction changes randomly with time, the wave is *randomly* polarized, or *unpolarized*. The vector **E** can also rotate uniformly at the wave frequency, as observed at a fixed point on the ray; the polarization is then *circular*, either right- or left-handed, depending on the direction of rotation.

Polarization plays an important part in the interaction of electromagnetic waves with matter, and Chapter 7 is devoted to a more detailed analysis.

1.6 The Electromagnetic Spectrum

The wavelength range of visible light covers about one octave of the electromagnetic spectrum, approximately from 400 to 800 nm (1 nanometre $= 10^{-9}$ m). The electromagnetic spectrum covers a vast range, stretching many decades through infrared light to radio waves and many more decades through ultraviolet light and X-rays to gamma rays (Figure 1.7). The differences in behaviour across the electromagnetic spectrum are very large. Frequencies (ν) and wavelengths (λ) are related to the velocity of light (c) by $\lambda\nu = c$. The frequencies vary from 10^4 Hz for long radio waves (1 hertz equals

[6] We use the SI system of electromagnetic units throughout.

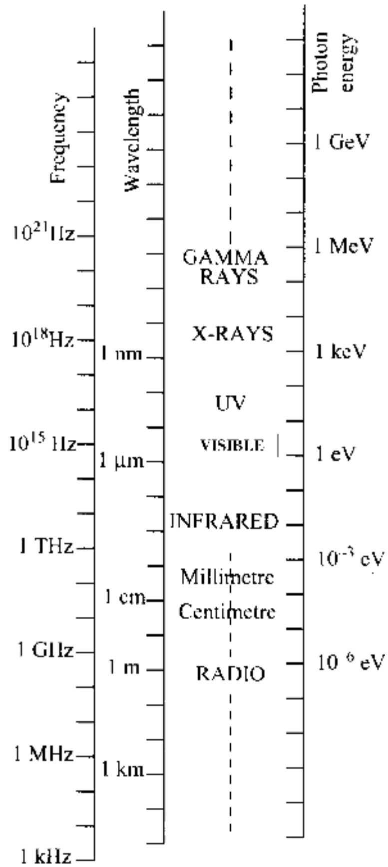

Figure 1.7 The electromagnetic spectrum

one cycle per second), to more than 10^{21} Hz for commonly encountered gamma rays; the highest energy cosmic gamma rays so far detected reach to 10^{35} Hz (4×10^{20} eV). It is unusual to encounter a quantum process in the radio frequency spectrum, and even more unusual to hear a physicist refer to the frequency of a gamma ray, instead of the energy and the momentum carried by a gamma ray photon.

Although wave aspects dominate the behaviour of the longest wavelengths, and photon aspects dominate the behaviour of short-wavelength X-rays and gamma rays, the whole range is governed by the same basic laws. It is in the optical range (waves in or near the visible range) that we most usually encounter the 'wave particle duality' which requires a familiarity with both concepts.

The propagation of light is determined by its wave nature, and its interaction with matter is determined by quantum physics. The relation of the energy of the photon to common levels of energy in matter determines the relative importance of the quantum at different parts of the spectrum: cosmic gamma rays, with a high photon energy and a high photon momentum, can act on matter explosively or like a high-velocity billiard ball, while long infrared or radio waves, with low photon energies, usually only interact with matter through classical electric and magnetic induction. We can explore these extremes in the following examples.

1. *What would be the velocity of a tennis ball, mass* 60 g, *with the same energy as a* 10^{20} eV *cosmic gamma ray photon*?

   Electron volt $= 1.602 \times 10^{-19}$ J. Kinetic energy $\frac{1}{2}mv^2 = 10^{20}\,\text{eV} = 10^{20} \times 1.6 \times 10^{-19}$ J. Velocity of 0.06 kg tennis ball is

$$v = \sqrt{\frac{2 \times 10^{20} \times 1.6 \times 10^{-19}}{6 \times 10^{-2}}} = 23\,\text{m}\,\text{s}^{-1} (= 83\,\text{km}\,\text{h}^{-1}).$$

2. *At what temperature would a molecule of hydrogen gas have, on average, the same energy as a photon of the* 21 *cm hydrogen spectral line?*

   In statistical physics each degree of freedom has an average energy of $\frac{1}{2}kT$. A hydrogen molecule has 5 degrees of freedom (3 translational and 2 rotational); hence thermal energy $= \frac{5}{2}kT$. Photon energy $h\nu = hc/\lambda$, so that $T = \frac{2}{5}hc/k\lambda = 0.068$ K.

3. *What wavelength of electromagnetic radiation has the same photon energy as an electron accelerated to* 100 *eV?*

   Photon energy $= h\nu = hc/\lambda = 100 \times 1.6 \times 10^{-19}$J. So

$$\lambda = \frac{6.63 \times 10^{-34} \times 3.00 \times 10^{8}}{1.6 \times 10^{-17}} = 1.24 \times 10^{-8}\,\text{m} = 12.4\,\text{nm}$$

   (ultraviolet light; see Figure (1.7)).

4. *An X-ray photon with wavelength* 1.5×10^{-11} *m arrives at a solid. How much energy (in eV) can it give to the solid?*

$$h\nu = \frac{hc}{\lambda} = \frac{6.63 \times 10^{-34} \times 3.00 \times 10^{8}}{1.5 \times 10^{-11}} = 1.32 \times 10^{-14}\,\text{J} = 8.3 \times 10^{4}\,\text{eV}.$$

The photon energy of visible light waves, ranging from 1.8 to 3.1 electron volts (eV), is such that quantum effects dominate only some of the processes of emission and absorption or detection. The visible spectrum contains the marks of quantum processes in the profusion of colour from line emission and in line absorption; it can also display a continuum of emission over a wide range of wavelengths, giving ‘white’ light, whose actual colour is determined by the large-scale structure of the continuum spectrum rather than its fine detail.

1.7 Stimulated Emission: The Laser

At the start of this chapter we remarked on the apparently complete understanding of optics at the beginning of the twentieth century. The wave nature of light was fully understood, stemming from the classical experiments of Young, Fresnel and Michelson, and substantiated by Maxwell’s electromagnetic theory. Much of the content of our later chapters on interference and diffraction is derived directly from that era (with some refinements). Even Planck’s bombshell announcement in 1900 that blackbody radiation is emitted by quantized oscillators, and Einstein’s demonstration in 1905 of the reality of photons through his explanation of the photoelectric effect, completed rather than disturbed the picture; they had cleared up a mystery about the interchange of energy between matter and

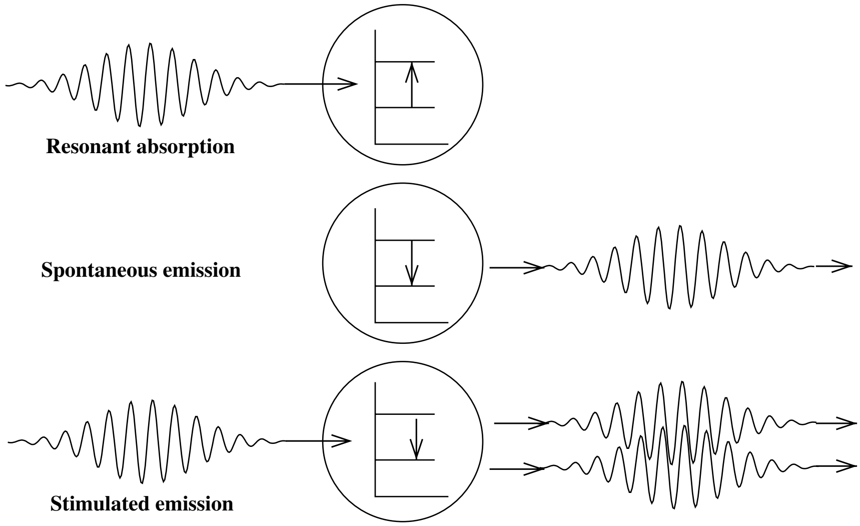

Figure 1.8 Three basic photon processes: absorption, spontaneous emission and stimulated emission. For simplicity only two energy levels are shown

electromagnetic waves. Einstein's theory of that interaction, however, contained the seed of another revolution in optics, which germinated half a century later with the invention of the laser.

Einstein in 1917 showed that there are three basic processes involved in the interchange of energy between a light wave and the discrete energy levels in an atom. All three involve a quantum jump of energy within the atom; typically in the visible region this is around 2 eV. Figure 1.8 illustrates the three basic photon processes; the processes are illustrated adopting a model with only two energy levels, although there are many more energy levels even in the simplest atom. As depicted in Figure 1.8, the first is the *absorption* of a photon which can occur when the quantum energy $h\nu$ of the photon equals the energy difference between the two levels (a *resonant* condition) and the photon falls on an atom in the lower level; the atom then gains a quantum of energy. The second is *spontaneous emission*, when an atom in the upper level emits a photon, losing a quantum of energy in the process. The third is *stimulated emission*, in which the emission of a photon is triggered by the arrival at an excited atom of another, resonant photon. This third process was shown by Einstein to be essential in the overall balance between emission and absorption. What emerged later was that the emitted photon is an exact copy of the incident photon, with the same direction, frequency and phase; further, each could then stimulate more photon emissions, leading to the build-up of a *coherent* wave which can attain a very great irradiance (or 'intensity', in old terminology).[7] The build-up requires the number of atoms in the higher energy level to exceed the number in the lower level, a condition known as population inversion, so that the rate of stimulated emission exceeds the rate of absorption. The energy supply used to create the population inversion is often referred to as a *pump*, which in Figure 1.9 is light absorbed between a ground level E_0 and level E_1. If the excitation of this level is short-lived, and it decays to a lower but longer-lived level E_2, the process leads to an accumulation

[7] See Appendix 1 for the definition of *irradiance* and other radiometric terms.

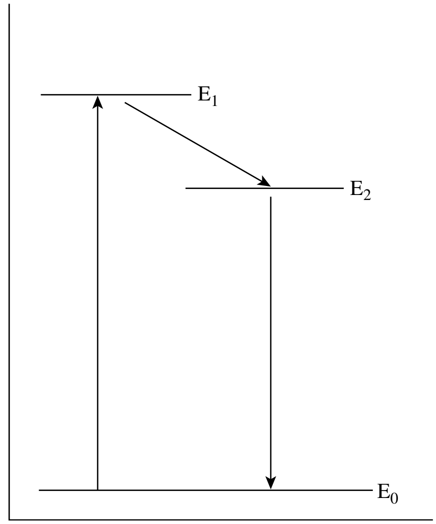

Figure 1.9 Energy levels in the three-level laser. Energy is supplied to the atom by absorption from the ground level to the excited level E_1; spontaneous emission to the long-lived level E_2 then results in overpopulation of that level. Transitions from E_2 to ground are then the stimulated emission in the laser

and overpopulation of atoms in the level E_2 compared with E_0. Stimulated emission, fed by energy from a pump, is the essential process in a *laser*. Prior to the laser, stimulated emission had been demonstrated in 1953 in the microwave region of the spectrum by Basov, Prokhorov and Townes,[8] an achievement for which they were awarded the Nobel Prize. We describe in Chapter 15 the earliest laser, due to T.H. Maiman in 1960.

The process of stimulated emission in a laser builds up a stream of identical photons, which add coherently as the most nearly ideal monochromatic light, with very narrow frequency spread and correspondingly great *coherence length* (Chapter 13). Paradoxically, lasers, which depend fundamentally on quantum processes, produce the most nearly ideal waves. Lasers have allowed the classical experimental techniques of interferometry and spectroscopy to be extended into new domains, which we explore in Chapter 9 on the measurement of length and Chapter 12 on high-resolution spectrometry.

Largely as a result of the discovery and development of lasers, a new subject of *photonics* has developed from pre-laser studies of transmission and absorption in dielectrics. Coherent laser beams easily achieve an irradiance many orders of magnitude greater than that of any thermal source, leading to very large electric fields and *non-linear* effects in dielectrics, such as harmonic generation and frequency conversion. There are many practical applications, some of which are more familiar in electronic communications, such as switching, modulation and frequency mixing. The title of this book indicates the current importance of lasers and photonics; the materials involved, including those used in non-linear optics, are included in Chapters 16 on laser light, 17 on semiconductors, 18 on light sources and 19 on detectors.

[8]They demonstrated a *maser* process, Microwave Amplification by the Stimulated Emission of Radiation. Note that strictly speaking this and the related laser process refer to amplification; devices which use the process in oscillators which generate microwaves and light are, however, known simply as masers and lasers.

1.8 Photons and Material Particles

As we noted in Section 1.1, the wave-like character of electrons was demonstrated in the 1920s, following the prediction by de Broglie that any particle with mass $m = E/c^2$ (where E is the total relativistic energy) and moving with velocity v has an associated wave with wavelength $\lambda = h/mv$. This association was eventually demonstrated in atoms, and even in molecules; in 1999 the wave–particle duality of the large molecule fullerene, or C_{60}, was demonstrated in a diffraction experiment by Arndt *et al.*[9]

There can be little doubt of the actual individual existence of a large particle such as a molecule of fullerene. Can we make a similar statement about the individual existence of photons? Ever since Planck and Einstein introduced quantum theory there has been a debate about the actual existence of photons as discrete objects. Light can be depicted as a ray, or as a wave; can it be thought of as a volley of photons, like a flock of birds moving from one roosting place to another? Should the wave nature of material particles, which constrains them to their behaviour in diffraction and interferometer observations, lead us to conclude that light has a similar dual nature?

Consider the classical interferometer typified by Young's double slit (Figure 1.10), which we describe in Chapter 8. Monochromatic light from the slit source passes through the pair of slits, forming an interference pattern on the screen. A detector on the screen records the arrival of individual photons, which in aggregate trace out the interference pattern, even when the intensity is so low that each recorded photon must have been the only photon present in the apparatus at any time. Through which slit did it pass? We naturally try to find out by placing some sort of detector at one or both slits, but as soon as we detect and locate the photon the interference pattern disappears. Detecting which slit the photon traverses has the same effect as forcing it to act like a localized quantum which passes through one slit at a time.

This behaviour is a simple example of the *complementarity principle* formulated by Bohr; if we know where the photon is, we cannot have an interference pattern, and if an interference pattern exists, it is impossible to specify the position of the photon. We can only observe that a photon has reached the detector, and the probability that it will arrive at any location is determined by its wave nature.

Diffraction and interference of material particles follow a similar pattern. In principle the double slit of Figure 1.10 could be demonstrating the de Broglie waves associated with a large molecule such as fullerene. Exactly the same dilemma arises: the interference pattern is observed even if only one molecule is in the apparatus at any time, but complementarity prevents us from knowing which slit the particle goes through, without destroying the interference pattern.

It has been suggested that the photon can exist in two places at once, and even that the large molecule is similarly 'delocalized'. This is better expressed by treating the wave as the basic description in both cases, and equating the probability of observing a particle or photon at a particular location to the intensity of the wave at that location. If any diffraction phenomenon is involved, the intensity pattern is determined by the *correlation* between separate wave components. If the separate components are 'de-correlated' by any process, the interference between wave components disappears. The analysis of correlation, which we present in Chapter 13, provides a unified framework for understanding diffraction both in light and in material particles. The difference, as noted in

[9] M. Arndt *et al.*, *Nature* **401**, 680, 1999.

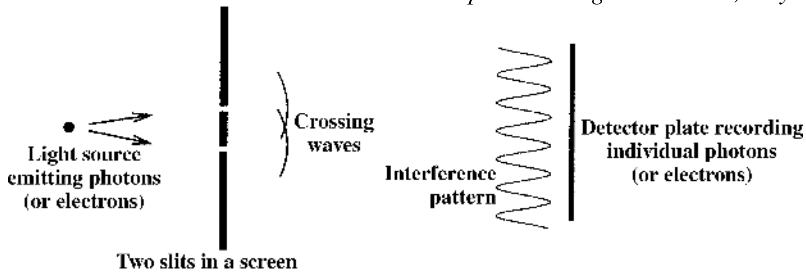

Figure 1.10 Double slit interferometer. Through which slit did each individual photon or electron go?

Section 1.1, is that a photon only exists as a quantized interchange between a field and an emitter or detector, while the individual existence of a material particle can hardly be questioned.

Problem 1.1

Gallium arsenide (GaAs) is an important semiconductor used in photoelectronic devices. It has a refractive index of 3.6. For a slab of GaAs of thickness 0.3 mm show that a point source of light within the GaAs on the bottom face will give rise to radiation outside the top face from within a circle of radius R centred immediately above the point source. Find R.

Problem 1.2

In the Pulfrich refractometer (Figure 1.11), the refractive index n of a liquid is found by measuring the emergent angle e from the prism whose refractive index is N. Show that if i is nearly 90°

$$n \approx (N^2 - \sin^2 e)^{1/2}.$$

Problem 1.3

The angular radius of a rainbow, measured from a point opposite to the Sun, may be found from the geometry of the ray in Figure 1.12, which lies in the meridian plane of a spherical drop of water with refractive index n. The

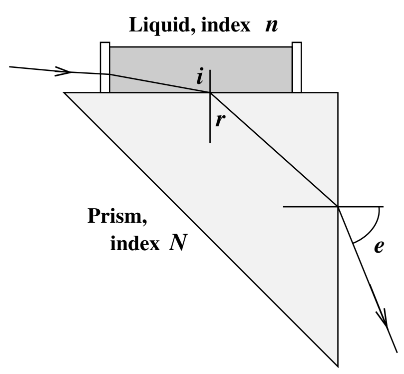

Figure 1.11 Pulfrich refractometer

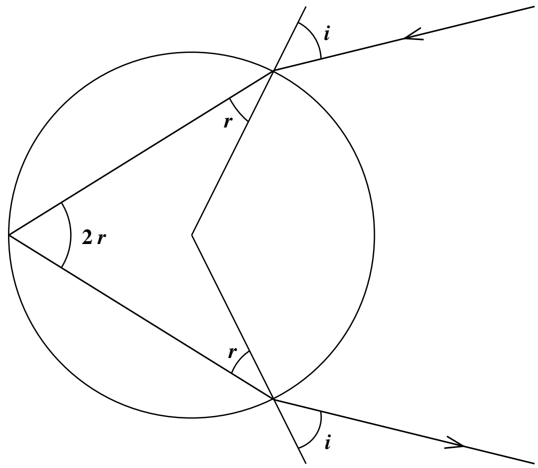

Figure 1.12 A ray refracted in the meridian plane of a spherical raindrop.

angular radius is a stationary value of the angle through which a ray from the Sun is deviated; show that it is given by

$$\cos i = \left(\frac{n^2 - 1}{3}\right)^{1/2}.$$

Note that the internal reflection is near the Brewster angle (see Section 5.4), so that the rainbow light is polarized along the circumference of the bow.

Problem 1.4
Show that the apparent diameter of the bore of a thick-walled glass capillary tube of refractive index n, as seen normally from the outside, is independent of the outer diameter, and is n times the actual diameter.

Problem 1.5
Show that the lateral displacement d of a ray passing through a plane-parallel plate of glass with refractive index n, thickness t, is related to the angle of incidence θ by

$$d \approx t\theta\left(1 - \frac{1}{n}\right)$$

provided that θ is small.

Problem 1.6
If the refractive index n of a slab of material varies in a direction y, perpendicular to the x axis, show by using Huygens' construction that a ray travelling nearly parallel to the x axis will follow an arc with radius

$$n\left(\frac{\mathrm{d}n}{\mathrm{d}y}\right)^{-1}.$$

(Consider a sector of wavefront δy across, and compare the distances travelled in time τ by secondary waves from each end of the sector.)

Problem 1.7

Show that the geometric distance of the horizon as seen by an observer at height h metres is approximately $3.5\,h^{1/2}$ kilometres. The radius of the Earth $\approx 6000\,\mathrm{km}$.

Use the result of Problem 1.6 to calculate how this is affected by atmospheric refraction, if this is due to pressure changes only with an exponential scale height of 10 kilometres. The refractive index of air at ground level is approximately 1.000 28.

Problem 1.8

The refractive index of solids at X-ray wavelengths is generally less than unity, so that a beam of X-rays incident at a glancing angle may be reflected, as in total internal reflection. If the refractive index is $n = 1 - \delta$ show that the largest glancing angle for reflection is $\simeq \sqrt{\delta}$. Evaluate this critical angle for silver at $\lambda = 0.07\,\mathrm{nm}$ where $\delta = 5.8 \times 10^{-6}$. (Note: if θ is the angle measured from the normal, the glancing angle is its complement, $\pi/2 - \theta$.)