OPTICS

W. T. Welford

W. T. WELFORD
PROFESSOR OF PHYSICS, IMPERIAL COLLEGE, LONDON

Optics

Oxford University Press · 1976

Oxford University Press, Ely House, London W.1

OXFORD LONDON GLASGOW NEW YORK
TORONTO MELBOURNE WELLINGTON CAPE TOWN
IBADAN NAIROBI DAR ES SALAAM LUSAKA ADDIS ABABA
KUALA LUMPUR SINGAPORE JAKARTA HONG KONG TOKYO
DELHI BOMBAY CALCUTTA MADRAS KARACHI

Casebound ISBN 0 19 851830 7
Paperback ISBN 0 19 851831 5

Printed in Great Britain
by Billing & Sons Limited, Guildford and Worcester

Preface

Present-day physics courses are under increasing pressure, on the one hand to keep up with developments in fundamental physics and on the other to cover a broad range of topics appropriate to the interests of students who may never become professional physicists. Thus the time available for optics in the first or second year of an undergraduate course, as for other branches of physics, decreases, and this has influenced my choice of topics in this book; I have been very selective and, as can be seen from the contents list, I have chosen material which is either basic to the development of the optics of the visible spectrum or which has interesting links with other kinds of optics or other branches of physics. Some may be concerned about what is *not* to be found in this book, e.g. measurement of the speed of light, group velocity, standing waves, the envelope function for diffr-action gratings, refractometry, Fresnel diffraction, and phase-change effects in interferometry. These omissions might have been dictated anyway by the agreed size of the Oxford Physics Series texts, but I do not plead this as an excuse. The book as it stands is intended as a reasonable selection of topics to be presented to undergraduates, perhaps in their first term at University and certainly having to cope with many other new things at the same time.

I have tried to stress physical arguments, and in order to reduce the mathematical complexity I have introduced the concept of a complex amplitude in the first chapter. I have also used the formalism of Fourier-transform theory freely, since this illuminates and simplifies every branch of physics in which waves appear; this may seem rather extreme for an elementary text, but since simple experiments with lasers are most easily discussed in terms of Fourier transforms it seems almost certain that students will meet the transform in their laboratory work and will grasp the basic ideas even if they have not been presented with a systematic formulation. However, sections 5.5 and 6.6 contain some more difficult Fourier-transform material, which could be omitted in very early courses. The main definitions and theorems of Fourier-transform theory needed are given, without proofs, in the Appendix.

Some of the problems at the end of each chapter amplify the text by introducing simple extensions of the main discussion.

I should like to thank my colleagues Dr. M.E. Barnett and Dr. R.W. Smith for their help with this book, mostly given unknowingly; many of their ideas about the teaching of optics have gone into it. Also I am very grateful to Professor E.J. Burge, who read the first draft, gave very valuable criticism, and made many useful suggestions, and to Miss Lesley Harwood, who prepared the index; and I thank the staff of Oxford University Press for their help during publication.

The quotations from James Joyce's *Ulysses* are by kind permission of the Society of Authors, as the literary representative of the Estate of James Joyce, and of The Bodley Head, as publishers.

Imperial College, W.T.W.
London, 1975

Contents

1. Waves, rays, and particles

But what I am anxious to arrive at is it is one thing to invent for instance those rays Röntgen did, or the telescope like Edison, though I believe it was before his time, Galileo was the man I mean. The same applies to the laws, for example, of a farreaching natural phenomenon such as electricity...

James Joyce, Ulysses

THE ELECTROMAGNETIC SPECTRUM

For many purposes optics can be regarded as the study of visible light, although in fact this light forms but a small part of a great range or spectrum of radiation. The most familiar part of this spectrum (apart from visible light) is probably the radio region (wireless waves). The complete spectrum of electromagnetic (e.m.) waves is described in Chapter 1 of *Radiation and quantum physics* (OPS 3) by D.J.E. Ingram. The waves are classified according to their wavelength λ or their frequency ν and these are related by

$$\lambda\nu = \text{velocity of the wave.} \tag{1.1}$$

Electromagnetic waves of all frequencies have the same velocity in vacuum, approximately 3.10^8 m s^{-1}; this universal constant is denoted by c.

We shall begin by describing light and other parts of the e.m. spectrum as electromagnetic waves, but this is only one possible description; light (as all other regions of the spectrum) has many properties which are better discussed in terms of other representations (e.g. rays or particles), and we shall have to consider these also.

An e.m. wave can be represented as in Fig.1.1. The graph represents the strength of the electric field in the wave at a given instant and at different points along the direction z of travel. Fig.1.2 shows the same thing in a more picturesque way;

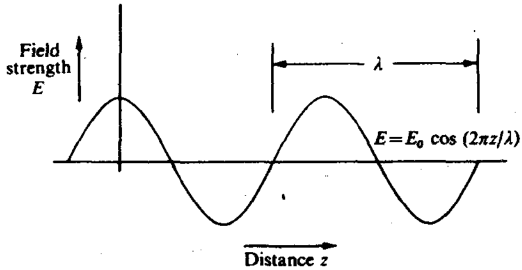

FIG.1.1. The electric field strength in an e.m. wave at a given instant as a function of the propagation distance z.

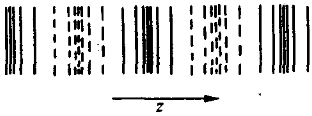

FIG.1.2. The amplitude of a wave. The closeness of the lines represents the field strength and broken lines indicate negative amplitudes.

the closeness of the lines indicates the relative strength of the electric field. Thus Figs 1.1 and 1.2 can be regarded as snapshots of the wave in space, taken at a certain instant of time. We could also look at a single point in space and consider the variation in time of the electric field at that point; we should then have a graph like Fig.1.3.

A more complete picture would be obtained by making the graph of Fig. 1.1 move along the z-axis at the velocity c of the wave. The field strength at any point as time passes would then vary as in Fig. 1.3. This travelling wave then has electric

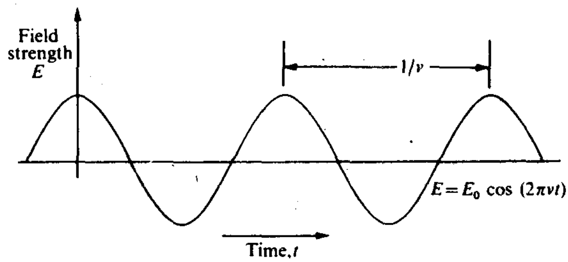

FIG. 1.3. The electric field strength in an e.m. wave at a given point as a function of time t

field strength E at any distance z and any time t given by

$$E = E_0 \cos 2\pi(\nu t - z/\lambda). \qquad (1.2)$$

This is easily verified by keeping t or z constant and comparing with the expressions in Figs 1.1 and 1.3 respectively.

To complete the picture of an e.m. wave we ought to consider also the accompanying magnetic field. But here it is sufficient to note that the magnetic field has a similar sinusoidal variation and that in the simplest situations, where the wave is not transferring energy to the medium through which it is travelling and where all parts of the wave are travelling in the same direction, the magnetic field varies in step or in phase with the electric field; both fields are at right-angles to the direction of travel of the wave.

Different sections of the e.m. spectrum are produced and detected in different ways, and the waves have a variety of interactions with matter, (see *Radiation and quantum physics* (OPS 3)). Although we shall be mainly concerned with visible light, it is easiest to consider first the properties of radio waves. This is because many of the properties we shall be interested in - those which produce interference and diffraction effects - can be demonstrated and explained for radio waves with fewer complications than for visible light.

POWER AND ENERGY

An essential property of all waves is that they transfer energy (from a source to a detector) without transferring the medium in which the waves occur. Indeed it is doubtful whether there can be said to be a "medium" for e.m. waves. Thus the rate of energy flow or the power in a wave is of interest. It follows from the detailed study of e.m. waves that for a wave like that in Figs 1.1 - 1.3 the power density (i.e. power per unit area across the wave transmitted in the direction of propagation) is proportional to the square of the electric field strength. We shall take this result as our starting point for a discussion of energy flow; it is treated in detail in texts on electromagnetism (e.g. *Electromagnetism* (OPS 1) by F.N.H. Robinson), where derivations and conditions of applicability can be found. Thus from (1.2) the power density is proportional to

$$E_0^2\{1 + \cos 4\pi(\nu t - z/\lambda)\}. \tag{1.3}$$

Clearly the cosine term causes a periodic fluctuation in energy flow across a certain plane, say $z = 0$. The oscillating electric field induces an alternating voltage in a conductor (antenna), and this constitutes detection of the e.m. wave.

One of the major differences between e.m. waves at radio and at optical (and higher) frequencies is that we have no detectors which can respond fast enough to demonstrate optical frequencies directly. In fact the fastest detectors of light will respond only to frequencies of the order of 10^9 - 10^{10} Hz, some 5 orders of magnitude too low. Thus any detector of e.m. radiation in the optical range responds only to the average power over many cycles of the waves. This *time-averaged power* is thus (from (1.3)) proportional simply to E_0^2.

THE COMPLEX EXPONENTIAL NOTATION AND THE COMPLEX AMPLITUDE

Another basic property of e.m. waves is that if two or more wave systems cross in a certain region of space, the electric and magnetic field strengths in this region are found simply by adding as vectors the fields from the individual wave systems. Thus we find the effect of overlapping waves by adding their field strengths or by linear combination. This simple result is not true for very large field strengths, and the topic of *nonlinear optics* has developed in the last decade now that such field strengths are available at optical frequencies. However, in this book we shall assume linear combination or *superposition*.

Both interference and diffraction phenomena can be explained in terms of superposition of waves, and in this section we shall discuss the mathematical symbolism for this.

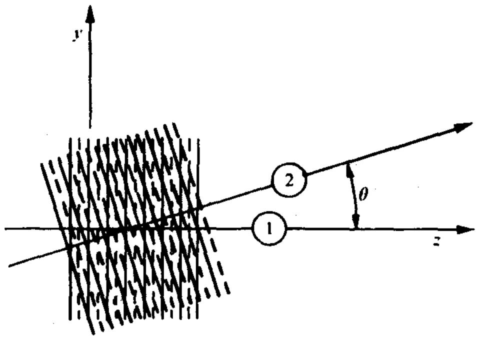

FIG. 1.4. The superposition of two e.m. waves travelling in directions at an angle θ to each other.

Suppose we have two e.m. waves of the kind described on p.1, travelling at an angle θ to each other, as in Fig. 1.4. Let the two waves have the same frequency (and therefore the same wave-length) and the same maximum field strength E_0. If we use axes as in the figure we can write the two waves as

$$\left.\begin{array}{l} E_0\cos 2\pi(\nu t - z/\lambda), \\ E_0\cos 2\pi\{\varepsilon + \nu t - (z\cos\theta + y\sin\theta)/\lambda\}. \end{array}\right\} \quad (1.4)$$

In the expression for the second wave the constant ε, known as a phase-shift term, allows for the possibility that the two waves are not in step at the origin of the coordinate system, and the expression $z\cos\theta + y\sin\theta$ ensures that the lines of constant electric field, or wavefronts, are at an angle θ to the y-axis. To fix our ideas we can regard each of the parallel lines in the figure as representing maximum field at a certain instant of time, but this is not essential. In order to find the *interference field*, as it is called, in the region where the waves cross we have to add the two expressions (1.4). If we are dealing with optical frequencies we can only observe the time-averaged power density, which is, of course, what we ordinarily know as the light intensity, and so we have to square the sum of the two expressions in (1.4) and find the time-average. There is no fundamental difficulty in doing this, but the manipulation of the trigonometrical expressions is very involved, particularly if we want to consider more than two waves and if they all have different field strengths. This has led to the introduction of the complex exponential notation and the use of the *complex amplitude* to describe waves, as follows.

First we replace an expression such as that in (1.2) by

$$E = E_0 \exp 2\pi i(\nu t - z/\lambda),$$

where i is, of course, $\sqrt{-1}$. We shall add complex expressions of this kind in superposing waves, but with the understanding that we are actually concerned only with the real parts. Since real and imaginary always remain separate in linear operations, this is valid. The above expression represents a wave with plane wavefronts travelling in the z-direction. We can now represent a similar plane wave, travelling in an arbitrary direction specified by a unit length vector $\underline{a}$, by the expression

$$E = E_0 \exp 2\pi i(\nu t - \underline{a}\cdot\underline{r}/\lambda),$$

where $\underline{r} = (x,y,x)$ is the vector from the origin to an arbitrary point in space. We can check that this agrees with the second of (1.4) by expanding the scalar product and remembering that the components of a unit vector are direction cosines.

Next we put $2\pi\nu = \omega$, the angular frequency, and we put $2\pi\underline{a}/\lambda = \underline{k}$. $\underline{k}$ is called the wave-vector, and we shall also use the scalar $|\underline{k}| = 2\pi/\lambda$, which we denote by k and call the wave-number. Thus our expression for a plane wave is

$$E(t,\underline{r}) = E_0 \exp i(\varepsilon + \omega t - \underline{k}\cdot\underline{r}). \qquad (1.5)$$

We have now indicated explicitly that the field strength E is a function of the position $\underline{r}$ and the time t, and we put in an arbitrary phase shift ε. We get the effect of superposing n waves of this kind by adding the appropriate terms,

$$\sum_n E_n \exp i(\varepsilon_n + \omega t - \underline{k}_n\cdot\underline{r}),$$

or, taking out the common factor $\exp i\omega t$, since we have supposed all the waves to have the same frequency,

$$\exp i\omega t \sum_n E_n \exp i(\varepsilon_n - \underline{k}_n\cdot\underline{r}).$$

We can write the summation, which is independent of the time, as $R + iI$, where R and I are two real functions of the position vector $\underline{r}$. From p.4 the intensity in the wave-field is the time-average of the square of the real part of

$$(R + iI)\exp i\omega t,$$

i.e, the time-average of

$$(R\cos\omega t - I\sin\omega t)^2.$$

It is easily verified that this time-average is simply $\frac{1}{2}(R^2 + I^2)$. The factor $\frac{1}{2}$ is usually dropped.

In this calculation the time-dependence of the waves appeared as a common factor $\exp i\omega t$ to all terms, which vanished in the final time-averaging; and the final intensity $R^2 + I^2$ is simply the squared modulus of the summed complex expressions.

Thus we have the rule that, to find the intensity due to several superposed plane waves of the same frequency, we add terms of the type $E_n\exp i(\varepsilon_n - \underline{k}_n\cdot\underline{r})$ for the individual waves and take the squared modulus at the end to find the intensity. An expression of the type

$$E\exp i(\varepsilon - \underline{k}\cdot\underline{r}),$$

in which the time-dependent part is omitted, is called a *complex amplitude*. These quantities can also be used for superposing other than plane waves (i.e. convergent or divergent waves), and for calculations with all forms of wave motion, not only e.m. waves. It is only necessary that the waves all have the same frequency. As a trivial example, the complex amplitude of the wave in eqn (1.2) is

$$E_0\exp(-2\pi iz/\lambda),$$

and the intensity is therefore immediately E_0^2. If we now apply the procedure to the two waves of eqn (1.4) we easily find, for the intensity in the plane $z = 0$, the expression

$$2E_0^2\,(1 + \cos\{(2\pi/\lambda)y\sin\theta\}).$$

This is a typical two-beam interference expression; we shall examine it more closely in Chapters 3 and 6.

As we noted earlier, the intensity, which has dimensions of power per unit area, is strictly proportional to E_0^2, i.e. in our present terms it is proportional to the squared modulus of the complex amplitude. The proportionality constant is important both for its dimensionality and for its numerical magnitude in connection with radio wave and microwave theory, but it is not important in the optical problems that we shall encounter. Thus for many purposes we can ignore the electromagnetic nature of light and discuss its properties in terms of a complex amplitude of some undefined quality or medium. Often we need not even specify whether the wave motion is transverse (e.m. waves or surface waves on water) or longitudinal (sound waves in air). This apparently abstract approach has advantages: parallels with other kinds of wave can be drawn, and we shall find it easier to come to terms with the fact that even the electromagnetic theory is not adequate to explain all optical phenomena.

It is found that all kinds of waves have to be characterized by two different quantities. These are of widely differing physical natures, depending on the kind of wave, but in all cases there is an *amplitude*, which varies in time and space and gives interference effects, and an *intensity*, which represents the rate of energy transport. With suitable interpretations the complex amplitude and its squared modulus, the intensity, can be used in all cases. All interference experiments and many diffraction experiments can be completely explained in these terms.

SOURCES AND DETECTORS

The production and detection of different parts of the e.m. spectrum are described in *Radiation and quantum physics* (OPS 3). Many of the effects and techniques which we usually call 'optical' apply mainly to the infrared, visible, and ultraviolet regions. In these regions there are three main kinds of source:

(1) thermal sources which produce a continuous spectrum, e.g. solid hot bodies, such as filament lamps, and hot gases under high pressure, as in an electric discharge through xenon (e.g. a flash tube);

(2) thermal sources giving line spectra, e.g. mercury vapour or neon discharge tubes under low pressure;

(3) lasers.

We can describe the production and detection of radio waves quite well in terms of the classical theory of electromagnetism, i.e. without invoking the existence of electrons or using quantum theory. However, in the optical region of the spectrum we have to introduce quantum concepts in order to explain light production and detection, although effects concerned with propagation alone (e.g. interference and diffraction) can be described in terms of a simple wave theory, usually involving only the use of the complex amplitude.

The quantum theory of light emission and absorption is explained in *Radiation and quantum physics* (OPS 3). Here we need only note that electromagnetic radiation is emitted or absorbed in finite quanta of energy called *photons*. The amount of energy in a photon depends on the frequency of the radiation and is given by

$$E = h\nu = hc/\lambda \qquad (1.6)$$

where h, the Planck constant, is 6.626×10^{-34} J s. The energy per photon is sometimes given in electronvolts (eV); 1 eV is 1.602×10^{-19} J. The emission or absorption of a photon corresponds to a change in the energy of an atom, molecule, or other system. In the infrared these transitions are between rotational or vibrational states of molecules; in the visible and near ultraviolet they correspond to changes in the energy levels of electrons in the outer orbits of an atom; and in the