Introduction to Applied Optics for Engineers

F. Paul Carlson

Introduction to Applied Optics for Engine

Introduction to Applied Optics for Engineers

F. Paul Carlson

Department of Electrical Engineering
University of Washington
Seattle, Washington

ACADEMIC PRESS New York San Francisco London 1977

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Carlson, F. Paul, Date
Introduction to applied optics for engineers.

Includes bibliographical references and index.
1. Optics. I. Title.
TA1520.C37 621.36 75-3575
ISBN 0-12-160050-5

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This book is designed to provide a broad base of material in applied optics for students and engineers who have a traditional background in electromagnetic wave theory. Often students are not exposed to the extension of electromagnetic wave concepts, to Kirchhoff-type integral expressions, or to the relation between wave concepts and optics. Traditionally, optics books assume an adequate background on the part of the reader in wave propagation, and minimize the presentation of the transitional material from simple wave concepts to integral equations. Thus a good understanding of the underlying wave theory of applied optics is not provided.

Once the basic integral expression describing the radiated fields is derived from Maxwell's equations, it is used to develop the laws and principles of optics and optical devices from the wave-picture point of view. This development enables the reader to understand the origin of the basic laws of optics, and enables the development of solutions for those cases where first-order approximations fail. The full-wave approach makes the presentation of diffraction theory much easier.

The book then goes on to describe the ideas of modern coherent optical data processing with examples taken from current research work. Several examples from bioengineering-related research are presented with suggestions for further work by the interested researcher.

Through this kind of development advanced seniors and first-year graduate students can obtain a grasp of the evolution and usefulness of optical devices such as lenses, wedges, arrays, and other processor elements. Design constraints and questions of physical realizability follow quite naturally from the limits imposed on specific mathematical approximations. This method of developing design limits from the approximation development limits the ambiguity associated with ad hoc developments.

With this early foundation centered on the Rayleigh Sommerfeld integral equation, the descriptions of optical processor systems, interferometers, lasers, holography, and other modern devices follow an evolutionary pattern of a mathematical description, and then go on to physical realization. In addition, the integral equation method is used as a basis for introducing and examining the ideas of partial coherence theory which are of current importance in the design of new detection devices and systems. Lastly, an overview of scattering theories is presented. The section on scattering theory begins with some simple observations of everyday phenomena and then proceeds to compare various current theories and research. The final section on multiple scattering provides a connection between the phenomenological and analytical approaches. Where possible, current and new results are presented to show how much optics, and particularly coherent optics, have developed in recent years. It seems that this is just a beginning in the quest to apply this technology.

This book is based on a set of course notes used in a one-quarter course in applied optics in the Department of Electrical Engineering at the University of Washington. In order to present this material in forty lectures it is necessary to keep a vigorous pace, particularly in the first four chapters. The problems are primarily designed to involve the student in understanding the text material, hence there are many derivations which are meant to be pedagogical. The style by which ideas are developed assumes that all the problems are worked. For some topics, the related ideas and developments occur in later chapters. This technique allows a smoother development of the basic ideas, and then provides a broader base for understanding actual systems and examples.

ACKNOWLEDGMENTS

I appreciate and acknowledge the comments and encouragement of Drs. Akira Ishimaru, David C. Auth, J. David Heywood, Sinclair Yee, Laurel J. Lewis, John L. Bjorkstam, James O. Raney and many others too numerous to list here. A special thank you is due to Mrs. E. Flewelling for her willingness and care in typing most of the manuscript. Finally, I am indebted most to my wife Judy, whose patience and encouragement was so helpful. The financial support of both the Office of Naval Research (Code 430C) and the National Institute of General Medical Sciences is acknowledged, with much appreciation, for much of the work reported on in the last six chapters.

CONTENTS

Chapter VII Interferometry

Chapter VIII Holography

Chapter 1

AN INTRODUCTION TO PHYSICAL OPTICS

Introduction

In this chapter the basic equations of physical optics are derived, starting with the differential form of Maxwell's equations. The wave treatment approach to optics leads quite nicely to the simpler geometrical (ray) optics approximation, and also shows the correspondence with the traditional plane-wave approach, which retains the important diffraction effects. This derivation enables the reader to understand the limits of the two approximations and use them appropriately.

Propagation in Free Space

To understand the physical observations encountered in radiating electromagnetic systems, such as optical systems, a convenient starting point is afforded through Maxwell's equations [1, 2]. These equations consist of four first-order differential equations. Together with the constitutive relations, a complete set is obtained, having a simple and tractable form. Solutions of these equations can be obtained when the boundary conditions are specified.

The method of solution using these four first-order equations is difficult because the variables are mixed. One approach to simplifying these equations is to eliminate the mixed-variable character by increasing the order. This procedure produces the wave equation which can be solved by simpler techniques.

The procedure starts with the time-dependent form of Maxwell's equations

$$\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t \tag{1-1a}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \partial D/\partial t \tag{1-1b}$$

$$\nabla \cdot \mathbf{D} = \varrho \tag{1-1c}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1-1d}$$

and the constitutive relations [1, 2]

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{1-2a}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{1-2b}$$

where ε and μ are the permittivity and permeability of the media. If we consider that the media in which the waves are propagating are isotropic and homogeneous, then the permittivity and permeability become simple scalar functions, and Eqs. (1-1) can be simplified. Furthermore, if the region is source free, then ϱ and $\mathbf{J}$ are zero, and Eqs. (1-1) become

$$\nabla \times \mathbf{E} = -\mu(\partial \mathbf{H}/\partial t) \tag{1-3a}$$

$$\nabla \times \mathbf{H} = \varepsilon(\partial \mathbf{E}/\partial t) \tag{1-3b}$$

$$\nabla \cdot \mathbf{E} = 0 \tag{1-3c}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1-3d}$$

Restricting consideration to only time-harmonic fields of the form $e^{-j\omega t}$ in the steady-state region, the time-differential terms further reduce to algebraic form

$$\nabla \times \mathbf{E} = j\omega\mu \mathbf{H} \tag{1-4a}$$

$$\nabla \times \mathbf{H} = -j\omega\varepsilon \mathbf{E} \tag{1-4b}$$

$$\nabla \cdot \mathbf{E} = 0 \tag{1-4c}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{1-4d}$$

where ω is the radian frequency of the waves or 2π times the frequency f in hertz.

Equations (1-4) still involve mixed variables and can be further reduced to a more tractable form by taking the curl either of Eq. (1-4a) or of (1-4b) and using the vector identity

$$\nabla \times \nabla \times \mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F} \tag{1-5}$$

To illustrate, the curl of Eq. (1-4a) is

$$\nabla \times \nabla \times \mathbf{E} = j\omega\mu(\nabla \times \mathbf{H}) \tag{1-6}$$

Substituting Eq. (1-4b) in the right-hand side leaves

$$\nabla \times \nabla \times \mathbf{E} = j\omega\mu(-j\omega\varepsilon\mathbf{E}) = \omega^2\mu\varepsilon\mathbf{E} \tag{1-7}$$

Since the divergence terms in Eqs. (1-4c) and (1-4d) are zero, Eq. (1-7) is simplified by the substitution of Eq. (1-5) to obtain

$$\nabla^2\mathbf{E} + \omega^2\mu\varepsilon\mathbf{E} = 0 \tag{1-8}$$

Equation (1-8) is known as the wave equation for the electric field. The derivation of a similar equation for the magnetic field **H** is included in the problem set at the end of this chapter. The vector form of Eq. (1-8) is more complicated than need be. In many cases a relation involving only a single component will suffice.

To obtain a complete solution to wave equation, appropriate boundary conditions are required corresponding to the physical structure. A common case is represented by zero tangential fields at perfect conductors and a radiation condition [2, 3] which requires the field to be zero at infinity. Basically, the wave equation describes how fields exist in time and space. The coupling through the two terms in the equation indicates that traveling or propagating waves will exist in the region. To illustrate, a plane-wave case will be considered.

Plane-Wave Propagation

Consider an infinite half space that has only the boundary condition defined by the radiation condition [2, 3], which arises from a constraint that finite energy exists at infinity. Using this condition, the solutions are simplified. In addition, consider that only a forward traveling wave exists. The explicit form of this assumption will follow in subsequent equations.

Throughout this treatment we have been considering macroscopic descriptions of the media by using ε and μ. Conventionally, however, the permittivity ε is replaced by the index of refraction η, which is

$$\eta = \sqrt{\varepsilon/\varepsilon_0} \tag{1-9}$$

where ε_0 is the permittivity of free space. The index of refraction basically enables a description of how the phase of a wave changes as a function of a relative or normalized permittivity ε_r, where

$$\varepsilon = \varepsilon_r \varepsilon_0 \tag{1-10}$$

One additional parameter is the notion of propagation vector **k**, where k^2 is defined as [1, 2]

$$k^2 = \omega^2 \mu \varepsilon \tag{1-11}$$

The vector direction is associated with the direction that the wave front propagates. In terms of free-space conditions, this can be written

$$k^2 = k_0{}^2 \eta^2 \tag{1-12}$$

Thus the wave equation of Eq. (1-8) reduces for one component to

$$(\nabla^2 + k_0{}^2 \eta^2) u = 0 \tag{1-13}$$

where u represents any of the three possible components E_x, E_y, or E_z. It should be noted that the Laplacian ∇^2 in the rectangular coordinates becomes

$$\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2 \tag{1-14}$$

If a simple one-dimensional problem is considered, Eq. (1-13) reduces to a simple form. The solution can be provided by an educated guess, or by noting that the form of the wave equation is like the classic harmonic function, leaving

$$u(x) = A e^{\pm jkx} \tag{1-15}$$

The amplitude of the wave is A, and $\pm jkx$ is the phase of the wave. Substitution of this result shows that it is indeed a solution. The plus/minus sign is used to indicate either forward or reverse waves. Remembering that an $e^{-j\omega t}$ time dependence was assumed, e^{+jkx} corresponds to a forward traveling wave.

The solution represented by Eq. (1-15) is known as the plane-wave solution. It is used extensively in the literature, primarily because most other complicated wave shapes can be decomposed into a sum of plane waves. It is called a plane wave because the phase fronts have no curvature or are said to be flat and constant in directions transverse to the direction of propagation. The wave vector **k** is also obtained quite simply from the gradient of kx, leaving $\hat{i}k$. Further, plane waves arise from sources that are very far away, thus satisfying the constraint of a source-free region. Because of the simple form associated with a plane-wave result, it will be used extensively in what follows.

Geometrical Optics

The plane-wave result, Eq. (1-15), suggests that a wave, satisfying the reduced scalar wave equation, can be represented by a simplified form having separable amplitude and phase terms of the form [3, 4]

$$u = Ae^{iS} \tag{1-16}$$

where A is the amplitude of the wave and S is the position-dependent phase term. Note that in the special case of Eq. (1-15), S is the scalar product of the propagation vector **k** and the position vector **x**. This representation suggests that the notion of phase fronts moving through space may adequately describe the field in a particular region. Noting that surfaces can generally be described by their normals, then surfaces of constant phase, with corresponding normals, may have significant meaning. Specifically, it is found that the normal is aligned with the general propagation vector **k**, as shown in Fig. 1-1.

In optics, particularly, these normals are referred to as rays. If the amplitude of the field A associated with a particular phase surface does not change appreciably over some appropriate distance of propagation, then the ray representation may embody most of the propagation phenomena, particularly in the short wavelength cases. That is, the wavelength is very

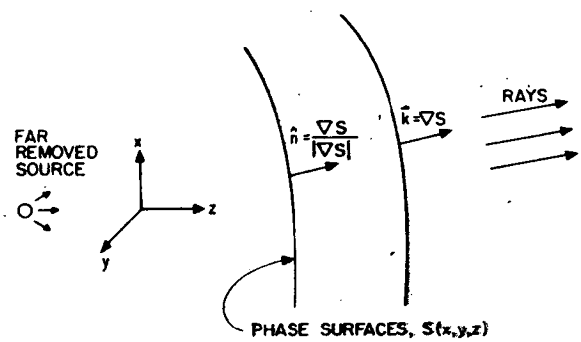

Fig. 1-1. Representation depicting equivalence of plane-wave models with large phase-front curvature and the ray-optic picture.

much smaller than the geometric constructs (e.g., lens sizes, focal lengths, turbulent eddy sizes) in the region. From this notion geometric optics can be defined as the limiting theory that describes the transport of energy in ray bundles. The easiest association with ray bundles is in the simple pictures describing lenses, wedges, and other ideas like Snell's law, which are first encountered in sophomore physics books. The notion of equivalence in these ideas is shown in Fig. 1-1.

In homogeneous media, these rays travel in straight lines, independent of each other. In inhomogeneous media, where the index of refraction is position dependent, the ray paths will be shown to be governed by a simple integral of the index of refraction over the traversed path. The basic law associated with this integral is Fermat's principle. One important result of this principle is that the paths may not necessarily be straight lines.

It can be shown how geometric optics follows from electromagnetic theory. To do so, use the plane-wave case governed by the homogeneous scalar wave equation, Eq. (1-13).

First, map $u(x, y, z)$ into the amplitude and phase space described by Eq. (1-16) [5]. Explicitly, divide Eq. (1-13) through by u and recognize that the first term can be related to the differential of $\log u$ as follows:

$$\nabla^2 u/u + k_0^2\eta^2 = \nabla^2(\log u) + (\nabla \log u)^2 + k_0^2\eta^2 = 0 \qquad (1\text{-}17)$$

If Eq. (1-16) is substituted, Eq. (1-17) becomes

$$\nabla^2(\log A) + (\nabla \log A)^2 - (\nabla S)^2 + k_0^2\eta^2 + i(\nabla^2 S + 2\nabla \log A \cdot \nabla S) = 0 \quad (1\text{-}18)$$

This complex equation can be separated into two independent equations by noting that the real and imaginary parts must each be zero. Thus the separated equations become

$$\nabla^2(\log A) + (\nabla \log A)^2 - (\nabla S)^2 + k_0^2\eta^2 = 0 \qquad (1\text{-}19\text{a})$$

and

$$\nabla^2 S + 2\nabla \log A \cdot \nabla S = 0 \qquad (1\text{-}19\text{b})$$

The mathematical implementation of the short-wavelength approximation follows by dropping certain terms using order of magnitude arguments. This will produce a shorthand equation that describes the case of geometric optics.

Without much loss of generality, the analysis can be restricted to the case where amplitude changes in the medium can occur only over distances on the order of scale size changes, such as a lens size, an inhomogeneity length, boundary size, etc. Denoting this scale by l, the short-wavelength approximation is represented by writing $l \gg \lambda$.

For the plane-wave representation, the gradient of S, which corresponds to the phase-surface normal, varies as the propagation vector $\mathbf{k}$. Further, Eq. (1-11), which defines the magnitude of $\mathbf{k}$, can be rewritten in terms of wavelength as

$$k = 2\pi/\lambda \tag{1-20}$$

Using this expression for k, the third term in Eq. (1-19a) can be written as

$$(\nabla S)^2 \sim (1/\lambda)^2 \tag{1-21}$$

The first two terms of Eq. (1-19a) can be reduced to

$$\nabla^2 A/A = \nabla^2(\log A) + (\nabla \log A)^2 \tag{1-22}$$

using the arguments leading to Eq. (1-17). However, since amplitude changes only occur over distances of the medium scale size, Eq. (1-22) varies like inverse length squared,

$$\nabla^2 A/A \sim (1/l)^2 \tag{1-23}$$

Since the last term in Eq. (1-19a) also varies like the inverse wavelength and since

$$(1/l)^2 \ll (1/\lambda)^2 \tag{1-24}$$

Eq. (1-19a) can be reduced to

$$(\nabla S)^2 = (k_0\eta)^2 \tag{1-25}$$

which is classically called the eiconal equation. Thus, the eiconal equation describes the phase-surface gradients or unnormalized surface normals in terms of the medium propagation vector. In this case the eiconal function S is the function describing the phase surface. The gradient of S leads to the notion of a ray traveling perpendicular to the surface [3].

Thus, solutions of the differential equation, Eq. (1-25), gave the wave fronts associated with a geometrical-optics representation of propagation. This is a useful concept for many cases. It has limitations, however, arising from the contributions of the dropped term $\nabla^2 A/A$. Physically, this term corresponds to the bending or curvature of the waves by medium objects. The description of the phenomenon of bending of waves around obstacles is given by diffraction theory.