FOUNDATIONS OF OPTICAL WAVEGUIDES

GILBERT H. OWYANG

Foundations of Optical Waveguides

Gilbert H. Owyang
Professor of Electrical Engineering
Worcester Polytechnic Institute

Edward Arnold

Published in the United Kingdom by
Edward Arnold (Publishers), Limited
41 Bedford Square
London WC1B 3DQ, England

First published in the United States of America
by Elsevier North Holland, Inc.
52 Vanderbilt Avenue, New York, New York 10017

ISBN 0 7131 3451 8

Desk Editor John Haber
Design Edmée Froment
Design Editor Virginia Kudlak
Art rendered by Vantage Art, Inc.
Cover Design by José García and Michele Maldeau
Production Manager Joanne Jay
Compositor Science Typographers, Inc.
Printer Haddon Craftsmen

Manufactured in the United States of America

Preface

As newer processes are developed for the manufacture of optical fibers, better quality is increasingly available while production costs quickly become more feasible. Consequently, communication in the optical frequencies is among the most rapidly developing areas of technology.

Although a number of books treating this material are presently available, they are generally designed for research scientists. This text is the result of notes developed for undergraduate and graduate courses in optical waveguides—including classes for practical engineers. It is intended to serve readers with some background in electromagnetic theory, but a thorough grounding is not assumed. Advanced undergraduates as well as practicing engineering scientists in the field of communications could find it useful. A basic understanding of wave propagation in bounded and unbounded regions should be sufficient prerequisite.

The text investigates from the field point of view the behavior of waves propagating in planar and cylindrical waveguides. Owing to the complexity of the problem, the analysis is mathematical in nature, but the physical interpretation of the theoretical results is emphasized throughout.

Beginning with a brief review of electromagnetic theory, wave propagation in free space, and guided waves in homogeneous media, the book analyzes the basic dielectric sheet, imperfect, and inhomogeneous waveguides. The more practical cladded cylindrical and inhomogeneous circular waveguides are also considered. Mathematical techniques involving eigenfunctions and Green's function are discussed in detail, as are methods of approximation.

One of the goals of this treatment is to enable readers to grasp fully the material in current research papers. Consequently, many techniques usually found only in higher-level mathematics textbooks are presented, eliminating the annoying and frustrating task of locating suitable references. While the text should remain accessible to those wishing to focus on the highlights of the mathematical development, it is intended to make further progress possible.

I would like to thank those reviewers whose criticisms and suggestions helped refine the manuscript. I also wish to extend my appreciation to John Haber of Elsevier North Holland, Inc., to Harit Majmudar, head of the Department of Electrical Engineering at Worcester Polytechnic Institute, for his support and encouragement, and to Geri Hicky for typing the manuscript.

Contents

1

Electromagnetic Theory

This chapter provides a survey of the background material that will be the basis of the entire book. The chapter begins with a review of the field relations and their mathematical solution. In order to analyze the wave equation, the method of separation of variables is introduced. The one-dimensional wave equation is investigated in detail and the results generalized to describe wave propagation in an arbitrary direction. The formulation of problems in unbounded nonhomogeneous media follows. Finally, wave propagation in a guiding structure filled with a homogeneous medium will be studied.

1.1. Maxwell's Equations

The *field* of a quantity is defined as the mathematical function that describes the variation of the quantity in a region. The electromagnetic field obeys Maxwell's equations,

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, t) = -\frac{\partial \tilde{\mathbf{B}}(\mathbf{r}, t)}{\partial t} \qquad \text{(Faraday's law)}, \tag{1}$$

$$\nabla \times \tilde{\mathbf{H}}(\mathbf{r}, t) = \tilde{\mathbf{J}}(\mathbf{r}, t) + \frac{\partial \tilde{\mathbf{D}}(\mathbf{r}, t)}{\partial t} \qquad \text{(Ampere's circuital law)}, \tag{2}$$

$$\nabla \cdot \tilde{\mathbf{B}}(\mathbf{r}, t) = 0, \tag{3}$$

$$\nabla \cdot \tilde{\mathbf{D}}(\mathbf{r}, t) = \tilde{\rho}(\mathbf{r}, t) \qquad \text{(Gauss's law)}, \tag{4}$$

where tildes label functions of position *and* time, and

$\tilde{\mathbf{E}}$ = electric field intensity in volts per meter,

$\tilde{\mathbf{H}}$ = magnetic field intensity in amperes per meter,

$\tilde{\mathbf{D}}$ = electric flux density in coulombs per square meter,

$\tilde{\mathbf{B}}$ = magnetic flux density in webers per square meter,

$\tilde{\mathbf{J}}$ = electric current density in amperes per square meter
$= \tilde{\mathfrak{J}}(\mathbf{r}, t) + \tilde{\mathbf{J}}_c(\mathbf{r}, t)$,

$\tilde{\mathfrak{J}}$ = the source current density,

$\tilde{\mathbf{J}}_c$ = the conduction current density,

$\tilde{\rho}$ = electric charge density in coulombs per cubic meter, and

$\mathbf{r}$ = the position vector of the field point.

Maxwell's equations are supplemented by the following *constitutive relations*, which characterize the properties of the medium:

$$\tilde{\mathbf{D}} = \epsilon \tilde{\mathbf{E}}, \tag{5}$$

$$\tilde{\mathbf{B}} = \mu \tilde{\mathbf{H}}, \tag{6}$$

$$\tilde{\mathbf{J}}_c = \sigma \tilde{\mathbf{E}}, \tag{7}$$

where

ϵ = the permittivity in farads per meter,

μ = the permeability in henry per meter, and

σ = the conductivity in mhos per meter.

Equations (5)–(7) are valid for linear isotropic media. A medium is linear if these relations hold independent of the magnitude of the field. A medium is isotropic if these relations hold independent of the direction of the field. The *constitutive parameters* (ϵ, μ, or σ) can be functions of position and time.

1.2. Periodic Time Dependent Fields

Maxwell's equations are partial differential equations in which the independent variables are the spatial coordinates and time. Consider the *simple-harmonic time varying field*, which can be expressed in the exponential form

$$\tilde{\mathbf{E}}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{j\omega t} \equiv \mathbf{E}e^{j\omega t}, \qquad \mathbf{E} \equiv \mathbf{E}(\mathbf{r}), \tag{1}$$

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r})e^{j\omega t} \equiv \mathbf{H}e^{j\omega t}, \qquad \mathbf{H} \equiv \mathbf{H}(\mathbf{r}). \tag{2}$$

Taking simple-harmonic time varying fields in the time domain does not restrict the applicability of the results. This may be readily understood when one recalls that any function of time can be represented as a Fourier series or Fourier integral of simple-harmonic functions. Once the solution of a simple-harmonic time varying field is known, the principle of superposition may be used to find the total field.

For simple-harmonic time varying fields, Maxwell's equations take the following form:

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}, \tag{3}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega\epsilon\mathbf{E}, \tag{4}$$

$$\nabla \cdot \mathbf{B} = \nabla \cdot (\mu\mathbf{H}) = \mathbf{H} \cdot \nabla\mu + \mu\nabla \cdot \mathbf{H} = 0, \tag{5}$$

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\epsilon\mathbf{E}) = \mathbf{E} \cdot \nabla\epsilon + \epsilon\nabla \cdot \mathbf{E} = \rho. \tag{6}$$

In a region excluding sources, that is, where $\mathcal{J}=0$ and $\rho=0$, Eqs. (4) and (6) become

$$\nabla \times \mathbf{H} = \mathbf{J}_c + j\omega\epsilon\mathbf{E} = (\sigma + j\omega\epsilon)\mathbf{E} = j\omega\varepsilon\mathbf{E}, \tag{7}$$

$$\varepsilon \equiv \epsilon(1 + \sigma/j\omega\epsilon), \tag{8}$$

$$\nabla \cdot \mathbf{D} = \mathbf{E} \cdot \nabla\epsilon + \epsilon\nabla \cdot \mathbf{E} = 0. \tag{9}$$

The complex permittivity, which will be written ε, is a convenient parameter, enabling Eq. (7) to have the same form whether the medium is a perfect conductor, a perfect insulator, or somewhere in between. Note that all field quantities in Eqs. (3)–(9) are functions of position only, the time dependency of these functions having been removed through Eqs. (1) and (2). The problem is thus simplified to finding the fields as function of position, that is, to solving the above equations. The complete expression of the fields as function of both position *and* time can then be obtained by factoring in $e^{j\omega t}$.

1.3. Generalized Wave Equations

Consider fields in a region which contains no sources, that is, where $\mathcal{J}=0$ and $\rho=0$. Maxwell's equations are then

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

$$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}, \qquad \varepsilon \equiv \epsilon(1 + \sigma/j\omega\epsilon). \tag{2}$$

The divergence of Eqs. (1) and (2) yields

$$\mathbf{H} \cdot \nabla\mu + \mu\nabla \cdot \mathbf{H} = 0, \tag{3}$$

$$\mathbf{E} \cdot \nabla\varepsilon + \varepsilon\nabla \cdot \mathbf{E} = 0. \tag{4}$$

These equations involve two field quantities, **E** and **H**, and can be reduced to a single equation for one field. The **H** field is first obtained from Eq. (1) and then substituted into Eq. (2):

$$\mathbf{H}=\frac{1}{-j\omega\mu}\nabla\times\mathbf{E},$$

$$\nabla\times\mathbf{H}=\nabla\times\left(\frac{1}{-j\omega\mu}\nabla\times\mathbf{E}\right), \tag{5}$$

$$=\nabla\left(\frac{1}{-j\omega\mu}\right)\times(\nabla\times\mathbf{E})+\frac{1}{-j\omega\mu}\nabla\times(\nabla\times\mathbf{E}),$$

$$=\frac{1}{j\omega\mu^2}\nabla\mu\times(\nabla\times\mathbf{E})+\frac{1}{-j\omega\mu}\nabla\times(\nabla\times\mathbf{E}).$$

Substitution of Eq. (5) into Eq. (2) yields

$$\nabla\times\nabla\times\mathbf{E}-\frac{1}{\mu}\nabla\mu\times(\nabla\times\mathbf{E})=-\gamma^2\mathbf{E}, \tag{6}$$

where

$$\gamma^2\equiv j\omega\mu\, j\omega\varepsilon=-\omega^2\mu\varepsilon \tag{7}$$

and γ is the *wave number* of the medium.

Similarly, if **E** is solved from Eq. (2) and then substituted into Eq. (1), one obtains

$$\nabla\times\nabla\times\mathbf{H}-\frac{1}{\varepsilon}\nabla\varepsilon\times(\nabla\times\mathbf{H})=-\gamma^2\mathbf{H}. \tag{8}$$

With use of the vector identity

$$\nabla\times\nabla\times\mathbf{g}=\nabla\nabla\cdot\mathbf{g}-\nabla^2\mathbf{g} \tag{9}$$

and Eq. (4), Eq. (6) then takes the following form:

$$\nabla\nabla\cdot\mathbf{E}-\nabla^2\mathbf{E}-\frac{1}{\mu}\nabla\mu\times(\nabla\times\mathbf{E})=-\gamma^2\mathbf{E},$$

$$\nabla\left(\frac{-1}{\varepsilon}\nabla\varepsilon\cdot\mathbf{E}\right)-\nabla^2\mathbf{E}-\frac{1}{\mu}\nabla\mu\times(\nabla\times\mathbf{E})=-\gamma^2\mathbf{E}, \tag{10}$$

$$\nabla^2\mathbf{E}+\nabla\left(\frac{1}{\varepsilon}\nabla\varepsilon\cdot\mathbf{E}\right)+\frac{1}{\mu}\nabla\mu\times(\nabla\times\mathbf{E})-\gamma^2\mathbf{E}=0.$$

Similarly, the use of Eqs. (3) and (9) in Eq. (8) yields

$$\nabla^2\mathbf{H}+\nabla\left(\frac{1}{\mu}\nabla\mu\cdot\mathbf{H}\right)+\frac{1}{\varepsilon}\nabla\varepsilon\times(\nabla\times\mathbf{H})-\gamma^2\mathbf{H}=0. \tag{11}$$

In the case where the constitutive parameters μ and ε are not functions of spatial variables, their gradients vanish and Eqs. (10) and (11) are simplified.

$$\nabla^2\mathbf{E}-\gamma^2\mathbf{E}=0, \tag{12}$$

$$\nabla^2\mathbf{H}-\gamma^2\mathbf{H}=0. \tag{13}$$

These are the *vector wave equations* and are valid for each field component in rectangular coordinates.

$$\nabla^2 E_i - \gamma^2 E_i = 0, \tag{14}$$

$$\nabla^2 H_i - \gamma^2 H_i = 0. \tag{15}$$

Here the subscript indicates the i component, where i stands for either x, y, or z. Equations (14) and (15) are known as the *scalar wave equations*, or *Helmholtz equations*.

1.4. One-Dimensional Wave Equation

It has been shown that both the electric and magnetic fields satisfy the wave equation,

$$\nabla^2 \mathbf{G} - \gamma^2 \mathbf{G} = 0, \qquad \gamma^2 \equiv -\omega^2 \mu \varepsilon, \tag{1}$$

where $\mathbf{G} \equiv \mathbf{G}(\mathbf{r})$ stands for either $\mathbf{E}$ or $\mathbf{H}$. When $\mathbf{G}$ is a function of a single variable, say z [i.e., $\mathbf{G} \equiv \mathbf{G}(z)$], Eq. (1) becomes

$$\frac{d^2 \mathbf{G}}{dz^2} - \gamma^2 \mathbf{G} = 0 \tag{2}$$

and each component of $\mathbf{G}$ satisfies the scalar wave equation

$$\frac{d^2 G_i}{dz^2} - \gamma^2 G_i = 0, \qquad i = x, y, \text{ or } z. \tag{3}$$

This is the *one-dimensional wave equation*. Let the trial solution of Eq. (3) be

$$G_i(z) = G_0 e^{pz}, \tag{4}$$

where G_0 is a constant and the quantity p is still to be determined. Then substitution of Eq. (4) into Eq. (3) yields

$$(p^2 - \gamma^2) G_0 e^{pz} = 0,$$

and therefore

$$p^2 - \gamma^2 = 0, \tag{5}$$

$$p = \pm \gamma. \tag{6}$$

Equation (5) is known as the *characteristic equation* of the problem, and its roots, given by Eq. (6), are the *characteristic values* (or *eigenvalues*) of the problem. The independent solutions corresponding to each characteristic value are known as the *characteristic functions* (or *eigenfunctions*) of the problem.

According to the theory of differential equations, the general solution of Eq. (3) is a linear combination of the eigenfunctions,

$$G_i(z) = G_i^+ e^{-\gamma z} + G_i^- e^{\gamma z}, \tag{7}$$

where G_i^+ and G_i^- are two arbitrary constants of integration for the second-order differential equation and can be determined by the specified boundary conditions. G_i^+ (G_i^-) is the amplitude of the positive (negative) traveling wave, as will be explained.

The sum of all three spatial components gives the complete vector expression of the field. In rectangular coordinates

$$\begin{aligned} \mathbf{G}(z) &= \hat{\mathbf{x}}(G_x^+ e^{-\gamma z} + G_x^- e^{\gamma z}) + \hat{\mathbf{y}}(G_y^+ e^{-\gamma z} + G_y^- e^{\gamma z}) \\ &\quad + \hat{\mathbf{z}}(G_z^+ e^{-\gamma z} + G_z^- e^{\gamma z}) \\ &= \mathbf{G}^+ e^{-\gamma z} + \mathbf{G}^- e^{\gamma z}, \end{aligned} \tag{8}$$

where $\mathbf{G}^{\pm} = \hat{\mathbf{x}}G^{\pm} + \hat{\mathbf{y}}G_y^{\pm} + \hat{\mathbf{z}}G_z^{\pm}$ and $\hat{\mathbf{x}}$ is a unit vector in the x direction. The complete expression of the field in spatial and time domains is then given by

$$\tilde{\mathbf{G}}(z,t) \equiv \mathbf{G}(z)e^{j\omega t} = \mathbf{G}^+ e^{j\omega t - \gamma z} + \mathbf{G}^- e^{j\omega t + \gamma z}. \tag{9}$$

The *wave number* or *propagation constant* γ is a complex quantity in general. From its definition, Eq. (1),

$$\begin{aligned} \gamma &= j\omega\sqrt{\mu\varepsilon} \\ &= j\omega\sqrt{\mu\epsilon(1+\sigma/j\omega\epsilon)} \equiv \alpha + j\beta. \end{aligned} \tag{10}$$

Then

$$e^{-\gamma z} = e^{-\alpha z} e^{-j\beta z}. \tag{11}$$

The real part of γ (and consequently $e^{-\alpha z}$) specifies the magnitude of the exponential function $e^{-\gamma z}$; α is therefore known as the *attenuation constant*, measured in nepers per meter. The factor $e^{-j\beta z}$ prescribes the phase of the function and consequently β is known as the *phase constant*, measured in radians per meter.

Equation (9) may be expressed

$$\mathbf{G} = G^+ e^{-\alpha z} e^{j(\omega t - \beta z)} + G^- e^{\alpha z} e^{j(\omega t + \beta z)}. \tag{12}$$

The above solution involves functions of two variables, $f(z,t) \equiv e^{j(\omega t \pm \beta z)}$. Such a function can be easily investigated by keeping one of the variables fixed and studying its variation with respect to the other variable. When $t=0$,

$$f(z, t=0) = e^{\pm j\beta z} \equiv f(z), \tag{13}$$

and when $t = t_1$, some arbitrary value,

$$f(z, t_1) = e^{j(\omega t_1 \pm \beta z)} = e^{\pm j\beta(z \pm \omega t_1/\beta)} \equiv e^{\pm j\beta z_1} \equiv f(z_1), \tag{14}$$

where $z_1 \equiv z \pm \omega t_1/\beta$. Since z and z_1 differ by a constant quantity ($\pm\omega t_1/\beta$), $f(z)$ and $f(z_1)$ must have the same shape. $f(z_1)$ is displaced from $f(z)$ by an amount ($\pm\omega t_1/\beta$) along the z axis. To be more specific, let

$$f^+(z)\equiv Ee^{j(\omega t-\beta z)}|_{t=0}=e^{-j\beta z}, \tag{15}$$

$$f^+(z_1)\equiv e^{j(\omega t-\beta z)}|_{t=t_1}=e^{-j\beta(z-\omega t_1/\beta)}=e^{-j\beta z_1}. \tag{16}$$

The relative location of $f^+(z)$ and $f^+(z_1)$ can be established at some reference point, say, the origin of each variable ($z=0$ and $z_1=0$).

$$z_1=0=z-\omega t_1/\beta \qquad \text{or} \qquad z_1=0 \quad \text{at} \quad z=\omega t_1/\beta. \tag{17}$$

Thus $f^+(z_1)$ is displaced by $\omega t_1/\beta$ in the positive z direction for positive values of t_1. Because $f^+(z,t)\equiv e^{j(\omega t-\beta z)}$ moves in the positive z direction as time increases, it is known as a *positive traveling wave function.*

Similarly, $f^-(z,t)\equiv e^{j(\omega t+\beta z)}$ is known as the *negative traveling wave function,* since the function moves in the negative z direction as time increases.

The speed at which the wave function travels,

$$e^{j\omega t\pm\gamma z}=e^{\pm\alpha z}e^{j(\omega t\pm\beta z)},$$

can be determined by observing the movement of a constant phase point. Suppose at some reference point $t=t_0$, $z=z_0$ the wave function has some phase of value M, so that

$$\omega t_0\pm\beta z_0=M. \tag{18}$$

At a later time $t_1=t_0+\Delta t$, this constant phase point will have traveled to a new position $z_1=z_0+\Delta z$. The phase is then given by

$$\begin{aligned}M&=\omega t_1\pm\beta z_1=\omega(t_0+\Delta t)\pm\beta(z_0\pm\Delta z)=(\omega t_0\pm\beta z_0)+\omega\Delta t\pm\beta\Delta z\\&=M+\omega\Delta t\pm\beta\Delta z\end{aligned}$$

or

$$\omega\Delta t\pm\beta\Delta z=0$$

$$\Delta z/\Delta t=\mp\omega/\beta. \tag{19}$$

The left-hand side of Eq. (19) has the dimension of velocity and can be interpreted as the average velocity of the constant phase point. In the above derivation, t_1 and z_1 are entirely arbitrary and they may be chosen as small (or large) as possible. If one takes the limit $\Delta t\to 0$, then

$$v_p=\lim_{\Delta t\to 0}\Delta z/\Delta t=\mathrm{d}z/\mathrm{d}t=\mp\omega/\beta. \tag{20}$$

One calls v_p the *phase velocity* of the traveling wave. The upper sign is for the traveling wave in the negative z direction, while the lower sign is for the positive traveling wave.

1.5. Method of Separation of Variables

The scalar wave equation for a scalar function $g(r)$ in rectangular coordinates is

$$\frac{\partial^2 g}{\partial x^2}+\frac{\partial^2 g}{\partial y^2}+\frac{\partial^2 g}{\partial z^2}-\gamma^2 g=0 \tag{1}$$

or

$$\frac{1}{g}\left(\frac{\partial^2 g}{\partial x^2}+\frac{\partial^2 g}{\partial y^2}+\frac{\partial^2 g}{\partial z^2}\right)-\gamma^2=0. \tag{2}$$

The method of separation of variables assumes a trial solution to be the product of three functions and each one is a function of a single variable only.

$$g(x,y,z)\equiv X(x)Y(y)Z(z). \tag{3}$$

Substitution of the trial solution into Eq. (2) yields

$$\frac{1}{X}\frac{d^2X}{dx^2}+\frac{1}{Y}\frac{d^2Y}{dy^2}+\frac{1}{Z}\frac{d^2Z}{dz^2}-\gamma^2=0. \tag{4}$$

Each of the first three terms can be a function of only one of the independent variables and the fourth term is a constant. Equation (4) requires the sum of all these terms to be constant, independent of all variables. This can be so if and only if each term is a constant itself. This can be verified by differentiating Eq. (4) with respect to one of the variables, say, x:

$$\frac{d}{dx}\left(\frac{1}{X}\frac{d^2X}{dx^2}\right)+\frac{d}{dx}\left(\frac{1}{Y}\frac{d^2Y}{dy^2}\right)+\frac{d}{dx}\left(\frac{1}{Z}\frac{d^2Z}{dz^2}\right)-\frac{d\gamma^2}{dx}=0$$

or

$$\frac{d}{dx}\left(\frac{1}{X}\frac{d^2X}{dx^2}\right)=0. \tag{5}$$

The quantity within the parentheses is thus a constant with respect to x. But by definition $X\equiv X(x)$, and therefore $(1/X)d^2X/dx^2$ is a constant, say, $\gamma_x{}^2$:

$$\frac{1}{X}\frac{d^2X}{dx^2}=\gamma_x{}^2 \qquad \text{or} \qquad \frac{d^2X}{dx^2}-\gamma_x{}^2X=0. \tag{6}$$