QUANTUM ELECTRONICS—PRINCIPLES AND APPLICATIONS

EDITED BY

PAUL F. LIAO
Bell Communications Research, Inc.
Red Bank, New Jersey

PAUL L. KELLEY
Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, Massachusetts

A complete list of titles in this series appears at the end of this volume.

THEORY OF DIELECTRIC OPTICAL WAVEGUIDES

Second Edition

Dietrich Marcuse

AT&T Bell Laboratories
Crawford Hill Laboratory
Holmdel, New Jersey

AT&T

Published by arrangement with AT&T

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Boston San Diego New York
London Sydney Tokyo Toronto

CONTENTS

Chapter 3. **Coupled Mode Theory**

Chapter 4. **Applications of the Coupled Mode Theory**

Chapter 5. **Coupled Power Theory**

Chapter 6. **Theory of the Directional Coupler**

Chapter 7. **Grating-Assisted Direction Couplers**

Chapter 8. **Approximate and Numerical Methods**

Chapter 9. **Nonlinear Effects**

Chapter 1

THE ASYMMETRIC SLAB WAVEGUIDE

1.1 Introduction

Dielectric slabs are the simplest optical waveguides. Because of their simple geometry, guided and radiation modes of slab waveguides can be described by simple mathematical expressions. The study of slab waveguides and their properties is thus often useful in gaining an understanding of the waveguiding properties of more complicated dielectric waveguides. However, slab waveguides are not only useful as models for more general types of optical waveguides, but they are actually employed for light guidance in integrated optics circuits [Mr1, Mr3].

A dielectric slab waveguide is shown schematically in Fig. 1.1.1. The figure shows a slab waveguide as it would be used in a typical integrated optics application. The core region of the waveguide is assumed to have refractive index n_1 and is deposited on a substrate with refractive index n_2. The refractive index of the medium above the core is indicated as n_3. The refractive index n_3 may be unity if the region above the core is air, or it may have some other

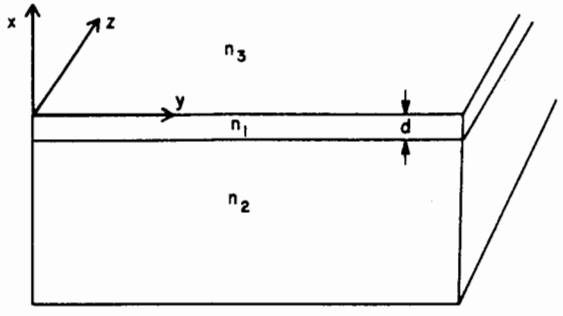

Fig. 1.1.1 *Schematic of a dielectric slab waveguide.*

value if the guiding region of index n_1 is surrounded by dielectric materials on both sides. In order to achieve true mode guidance it is necessary that n_1 be larger than n_2 and n_3. In order to have a specific example in mind we shall assume that

$$n_1 > n_2 \geqslant n_3 \tag{1.1-1}$$

If $n_2 = n_3$, we speak of a symmetric slab waveguide. In case that $n_2 \neq n_3$, the slab waveguide is asymmetric. The modes of symmetric slab waveguides are simpler than those of asymmetric slabs because they can be expressed either as even or odd field distributions [Mel]. The lowest-order mode of a symmetric slab waveguide does not have a cutoff frequency, which means that, in principle, this mode can propagate at arbitrarily low frequencies. By contrast, all modes of asymmetric slabs become cutoff if the frequency of operation is sufficiently low.

This chapter is devoted to the description of the optical properties of asymmetric slab waveguides. Since symmetric slabs are only limiting cases of asymmetric slabs, the description of symmetric slab waveguides is necessarily included in our present treatment.

Like all dielectric waveguides the asymmetric slab supports a finite number of guided modes which is supplemented by an infinite continuum of unguided radiation modes. Both types of modes are obtained as solutions of a boundary value problem. However, the guided modes can also be considered from the point of view of ray optics. Since ray optics is intuitively more appealing than wave optics, we start the discussion by deriving the eigenvalue equation of the guided modes from geometrical optics, which is supplemented by some simple results of plane wave reflection and refraction at plane dielectric interfaces.

The boundary value problem is discussed and solved in Sects. 1.3 and 1.4 on wave optics of the guided modes of the slab waveguide. The chapter then continues with a discussion of leaky waves and of an inverted slab waveguide with $n_1 < n_2, n_3$ which supports only leaky waves.

In integrated optics applications, slab waveguides are formed by various means, the simplest of which use the deposition of glass or plastic films on glass or plastic substrates. These films can be deposited by evaporation, sputtering, or by epitaxial growth techniques. The last method is restricted to the deposition of thin single crystalline films on crystal substrates. Another method of forming dielectric optical waveguides for integrated optics applications employs ion implantation techniques. By bombarding the substrate material with suitable ions it is possible to alter the refractive index of the substrate so that a dielectric slab waveguide results. The depth at which the guiding region appears below the substrate surface can be controlled by the choice of the energy that is used to accelerate the ion beam.

Many integrated optics applications use narrow dielectric strip waveguides instead of a continuous two-dimensional film. The modes of such structures are discussed in Sect. 1.7. Such waveguides are formed by ion implantation techniques or by the deposition of a thin film on top of a substrate which is subsequently etched away, so that only the narrow strip waveguides are left.

The study of asymmetric slab waveguides serves as a valuable introduction to the entire field of dielectric optical waveguides. Because of their simplicity, slab waveguides provide insight into the mechanism of waveguidance by dielectric optical waveguides.

1.2 Geometrical Optics Treatment of Slab Waveguides

Geometrical (or ray) optics describes the propagation of light fields by defining rays as the lines that cross the surfaces of constant phase of the light field at right angles. Light rays have intuitive appeal since a narrow beam of light is a good approximation to the more abstract notion of light rays.

The laws of ray optics, needed for our present purpose, are simple. We need only assume that a light ray in a homogeneous optical medium follows a straight path. In addition, we need to know Snell's law, which relates the angles with respect to the normal to a dielectric interface that a beam forms that passes through this interface. With the definition of the angles shown in Fig. 1.2.1 Snell's law can be expressed in the form [Mel]

$$n_1 \sin\alpha_1 = n_2 \sin\alpha_2 \tag{1.2-1}$$

For our purposes it is more convenient to use the angle between the ray and the dielectric interface, so that Snell's law assumes the form

$$n_1 \cos\theta_1 = n_2 \cos\theta_2 \tag{1.2-2}$$

If $n_1 > n_2$ it is apparent from Eq. (1.2-2) that there is no real angle θ_2 if $n_1 \cos\theta_1 > n_2$. The absence of a real angle satisfying Eq. (1.2-2) can be interpreted as total internal reflection. In this case no light beam emerges on the

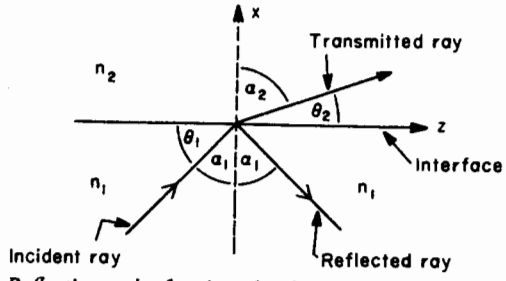

Fig. 1.2.1 *Reflection and refraction of a plane wave at a dielectric interface.*

opposite side of the dielectric interface, so that all the light is totally reflected inside medium 1. The critical angle for total internal reflection is defined by

$$\cos\theta_{1c} = n_2/n_1 \tag{1.2-3}$$

A straight light ray can be associated with a plane wave. Ray optics does not include the description of the phase of the light field. However, the notion of the optical path length, which is defined by the actual length times the refractive index of the medium, makes it easy to attach a phase to the light ray. A ray of length s has traveled an optical path length $n_1 s$ in medium 1. Its phase relative to the starting point at $s = 0$ can be defined as $\phi = -n_1 ks$ (the minus sign is necessary since we assume time dependence $e^{i\omega t}$) with the free-space propagation constant of plane waves,

$$k = 2\pi/\lambda \tag{1.2-4}$$

However, a plane wave or a ray accumulates phase shift not only by traveling in space but also by reflection from a dielectric interface. The reflection coefficient of a plane wave at a dielectric interface, which is polarized so that its electric vector is parallel to the interface, follows from Eqs. (1.6-10), (1.6-11), and (1.6-14) of [Mel]:

$$r = \frac{B}{A} = \frac{(n_1^2k^2-\beta^2)^{1/2} - (n_2^2k^2-\beta^2)^{1/2}}{(n_1^2k^2-\beta^2)^{1/2} + (n_2^2k^2-\beta^2)^{1/2}} \tag{1.2-5}$$

The parameter β (which was called k_{1z} in [Mel]) can be expressed in terms of the angle θ_1:

$$\beta = n_1 k \cos\theta_1 \tag{1.2-6}$$

For $\theta_1 > \theta_{1c}$, the reflection coefficient r, given by Eq. (1.2-5), is real and positive, so that no additional phase change occurs on reflection from the medium with index n_2. For $\theta_1 < \theta_{1c}$, total internal reflection occurs at the dielectric interface. In this case r is complex because the second square root in numerator and denominator of Eq. (1.2-5) is negative imaginary. The

negative sign is necessary since a decaying instead of a growing wave must result in medium 2. Under conditions of total internal reflection, a wave that is polarized with its electric vector parallel to the interface suffers a phase shift

$$\phi = -2\arctan[(\beta^2-n_2^2k^2)^{1/2}/(n_1^2k^2-\beta^2)^{1/2}] \tag{1.2-7}$$

For a wave polarized so that its magnetic vector is parallel to the interface, we obtain from Eq. (1.6-56) of [Mel] the phase shift

$$\phi = -2\arctan[(n_1^2/n_2^2)(\beta^2-n_2^2k^2)^{1/2}/(n_1^2k^2-\beta^2)^{1/2}] \tag{1.2-8}$$

Having collected these few facts from ray optics and the theory of plane wave reflection at dielectric interfaces enables us to discuss mode guidance in the slab waveguide and derive the eigenvalue equation for the propagation constants of the guided modes. The rays, which belong to the mode field, propagate inside the core region with refractive index n_1. They are confined to this region by suffering total internal reflection at the dielectric interfaces. Figure 1.2.2 shows the trajectory of such a ray.

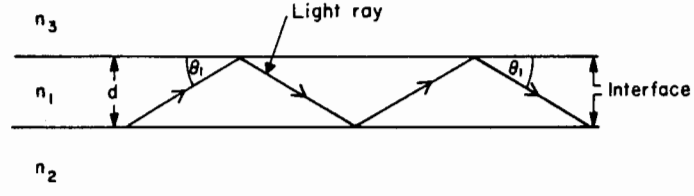

Fig. 1.2.2 *Ray trajectory of a guided wave in the slab waveguide.*

So far we have accounted for the mechanism of mode confinement and have indicated that the angle θ_1 must not exceed the critical angle of Eq. (1.2-3). One might reach the conclusion that rays with arbitrary angles can propagate in the slab waveguide provided the critical angle is not exceeded. However, this is an erroneous assumption. In order to understand the reason for the occurrence of discrete angles for rays that are associated with guided modes, we must introduce our knowledge of the phase of the ray into the picture. The rays shown in Fig. 1.2.2 are only a convenient description of the plane waves that are associated with them. Figure 1.2.3 shows the phase fronts of the plane waves as dashed lines. Two points, A and B, are marked in the figure. The phase fronts that go through these two points both belong to the same plane wave. The phase fronts of the reflected wave traveling downward have been omitted in order to keep the picture as simple as possible. The ray from point A to point B (ray AB) is assumed to have suffered no reflection. The longer ray from point C to point D (ray CD) belonging to the reflected wave has

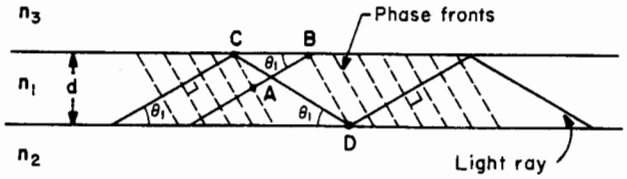

Fig. 1.2.3 *Illustration of the phase condition that leads to the eigenvalue equation. All rays that travel in the same direction belong to the same plane wave.*

suffered two total internal reflections as it travels from the phase front through A to the phase front through point B. Since all points on the same phase front of a plane wave must be in phase, we must require that the optical path length of the ray AB differ from that of the ray CD by a multiple of 2π. The distance between points B and C is $(d/\tan\theta_1)-d\tan\theta_1$. The distance between points A and B is, therefore,

$$s_1 = [(1/\tan\theta_1)-\tan\theta_1]d\cos\theta_1 = (\cos^2\theta_1-\sin^2\theta_1)d/\sin\theta_1 \tag{1.2-9}$$

The length of the long ray through points C and D is

$$s_2 = d/\sin\theta_1 \tag{1.2-10}$$

However, in addition to the phase change accumulated in traveling the actual distance s_2, ray CD has suffered two total internal reflections which resulted in the phase change ϕ_3 (reflection from the interface with region 3) and ϕ_2 (reflection from the interface with region 2). The condition that both rays contribute to the same plane waves can thus be expressed by the relation

$$n_1(s_2-s_1)k + \phi_2 + \phi_3 = 2N\pi \tag{1.2-11}$$

where N is an integer number. Equation (1.2-11) is a condition that determines the allowed ray angles or, through relation (1.2-6), the value of the parameter (eigenvalue) β. Such a relation is known as an eigenvalue equation. Since the phase shifts ϕ_2 and ϕ_3 are different for the two possible polarizations of the waves, we obtain two different eigenvalue equations. We designate the waves with electric vectors parallel to the interface as TE waves and obtain from Eqs. (1.2-7), (1.2-9), (1.2-10), and (1.2-11) the eigenvalue equation

$$\arctan(\gamma/\kappa) + \arctan(\delta/\kappa) \mp \kappa d - N\pi \tag{1.2-12}$$

with the abbreviations

$$\kappa = (n_1^2k^2-\beta^2)^{1/2} = n_1 k\sin\theta_1 \tag{1.2-13}$$

$$\gamma = (\beta^2-n_2^2k^2)^{1/2} = [(n_1^2-n_2^2)k^2-\kappa^2]^{1/2} \tag{1.2-14}$$

and

$$\delta = (\beta^2-n_3^2k^2)^{1/2} = [(n_1^2-n_3^2)k^2-\kappa^2]^{1/2} \tag{1.2-15}$$

Taking the tangent of Eq. (1.2-12) transforms the eigenvalue equation for the TE waves into the form

$$\tan\kappa d = \kappa(\gamma+\delta)/(\kappa^2-\gamma\delta) \tag{1.2-16}$$

The corresponding procedure for TM waves results in the eigenvalue equation

$$\tan\kappa d = n_1^2\kappa(n_3^2\gamma + n_2^2\delta)/(n_2^2n_3^2\kappa^2 - n_1^4\gamma\delta) \tag{1.2-17}$$

We shall rederive these equations again in the next section by starting from Maxwell's equations and using the boundary conditions at the dielectric interfaces. The purpose of the present discussion was to give an intuitive explanation for the process of mode guidance in slab waveguides and to demonstrate how the mode conditions (1.2-16) and (1.2-17) can be obtained with the help of simple principles obtained from ray optics and from the properties of plane waves [Ti1].

1.3 Guided Modes of the Asymmetric Slab Waveguide

In order to obtain a complete description of the modes of dielectric waveguides, Maxwell's equations must be solved [MK1, NM1]. With the help of the operator ($\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are unit vectors in x, y, z direction)

$$\nabla = \mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z\frac{\partial}{\partial z} \tag{1.3-1}$$

Maxwell's equations can be written in the form

$$\nabla \times \mathbf{H} = \varepsilon_0 n^2\,\partial\mathbf{E}/\partial t \tag{1.3-2}$$

and

$$\nabla \times \mathbf{E} = -\mu_0\,\partial\mathbf{H}/\partial t \tag{1.3-3}$$

The cross indicates a vector product, $\mathbf{H}$ and $\mathbf{E}$ are the magnetic and electric field vectors, and ε_0 and μ_0 are the dielectric permittivity and magnetic permeability of vacuum. We do not consider magnetic materials in this book so that the use of the vacuum constant μ_0 is sufficient. The index of refraction of the medium is designated by n, and t is the time variable.

We simplify the description of the slab waveguide by assuming that there is no variation in y direction, which we express symbolically by the equation

$$\partial/\partial y = 0 \tag{1.3-4}$$

Condition (1.3-4) is actually no restriction on the generality of the mode description since it is always possible to rotate the coordinate system in the yz plane until this condition is satisfied for any given mode. The modes of the slab waveguide can be classified as TE and TM modes. TE or transverse

electric modes do not have a component of the electric field in the direction of wave propagation, while TM or transverse magnetic modes do not have a longitudinal magnetic field component. We consider TE and TM modes separately. The fields of guided modes must vanish at $x = \pm\infty$.

Guided TE Modes

TE modes have only three field components: E_y, H_x, and H_z. The position of the coordinate system relative to the slab is shown in Fig. 1.1.1. We assume that the slab is infinitely extended in the yz plane. We consider only strictly time harmonic fields whose time dependence, in complex notation, can be expressed as

$$e^{i\omega t} \tag{1.3-5}$$

The radian frequency ω is related to the actual frequency f by

$$\omega = 2\pi f \tag{1.3-6}$$

Since we are interested in obtaining the normal modes of the slab waveguide we assume that the z dependence of the mode fields is given by the function

$$e^{-i\beta z} \tag{1.3-7}$$

By combining the two factors (1.3-5) and (1.3-7) we obtain

$$e^{i(\omega t - \beta z)} \tag{1.3-8}$$

Function (1.3-8) describes a wave traveling in positive z direction with phase velocity

$$v = \omega/\beta \tag{1.3-9}$$

The eigenvalue β is identical to the quantity introduced in Eq. (1.2-6). Factor (1.3-8) is common to all field quantities and shall be omitted for brevity. With $E_x = 0$, $E_z = 0$, and $H_y = 0$, we obtain from Maxwell's Eqs. (1.3-2), (1.3-3) with the help of Eq. (1.3-4), and with the time and z dependence given by Eq. (1.3-8),

$$-i\beta H_x - (\partial H_z/\partial x) = i\omega\varepsilon_0 n^2 E_y \tag{1.3-10}$$

$$i\beta E_y = -i\omega\mu_0 H_x \tag{1.3-11}$$

$$\partial E_y/\partial x = -i\omega\mu_0 H_z \tag{1.3-12}$$

We thus obtain the H components in terms of the E_y component

$$H_x = (-i/\omega\mu_0)\,\partial E_y/\partial z = -(\beta/\omega\mu_0)\,E_y \tag{1.3-13}$$

and

$$H_z = (i/\omega\mu_0)\,\partial E_y/\partial x \tag{1.3-14}$$

Substitution of these two equations into Eq. (1.3-10) yields the one-dimensional reduced wave equation for the E_y component

$$(\partial^2 E_y/\partial x^2) + (n^2k^2 - \beta^2)\,E_y = 0 \tag{1.3-15}$$

with $k^2 = \omega^2\varepsilon_0\mu_0 = (2\pi/\lambda)^2$. The problem of finding the TE modes of the slab waveguide has thus become very simple. We only need to find solutions of the one-dimensional reduced wave Eq. (1.3-15) and obtain the magnetic field components directly from Eqs. (1.3-13) and (1.3-14). The only remaining complication is the requirement that the solutions must satisfy the boundary conditions at the two dielectric interfaces at $x = 0$ and $x = -d$. The boundary conditions require that the tangential E and H fields be continuous at the dielectric discontinuities. We thus must require that E_y and H_z are continuous at $x = 0$ and $x = -d$. Solutions that satisfy these conditions for the E_y component and vanish at $x = \pm\infty$ are

$$E_y = Ae^{-\delta x}, \qquad \text{for} \quad x \geqslant 0 \tag{1.3-16}$$

$$= A\cos\kappa x + B\sin\kappa x, \qquad \text{for} \quad 0 \geqslant x \geqslant -d \tag{1.3-17}$$

$$= (A\cos\kappa d - B\sin\kappa d)\,e^{\gamma(x+d)}, \qquad \text{for} \quad x \leqslant -d \tag{1.3-18}$$

The refractive index n in Eq. (1.3-15) assumes the value n_3 in region 3 (see Fig. 1.1.1), for $x > 0$; n_1 in region 1, for $0 > x > -d$; and n_2 in region 2, for $x < -d$. The parameters κ, γ, and δ are defined by Eqs. (1.2-13)–(1.2-15). The E_y component shown by these three equations satisfies the reduced wave Eq. (1.3-15) and is continuous at the two dielectric interfaces. We do not need the H_x component for the moment. The H_z component is obtained from Eq. (1.3-14):

$$H_z = (-i\delta/\omega\mu_0)\,Ae^{-\delta x}, \qquad \text{for} \quad x \geqslant 0 \tag{1.3-19}$$

$$= (-i\kappa/\omega\mu_0)(A\sin\kappa x - B\cos\kappa x), \qquad \text{for} \quad 0 \geqslant x \geqslant -d \tag{1.3-20}$$

$$= (i\gamma/\omega\mu_0)(A\cos\kappa d - B\sin\kappa d)\,e^{\gamma(x+d)}, \qquad \text{for} \quad x \leqslant -d \tag{1.3-21}$$

The H_z component does not immediately satisfy the boundary conditions. The requirement of continuity of H_z at $x = 0$ and $x = -d$ leads to the following system of equations:

$$\delta A + \kappa B = 0 \tag{1.3-22}$$

$$(\kappa\sin\kappa d - \gamma\cos\kappa d)\,A + (\kappa\cos\kappa d + \gamma\sin\kappa d)\,B = 0 \tag{1.3-23}$$

This homogeneous equation system has a solution only if the system determinant vanishes. We thus obtain the determinantal or eigenvalue equation

$$\delta(\kappa\cos\kappa d + \gamma\sin\kappa d) - \kappa(\kappa\sin\kappa d - \gamma\cos\kappa d) = 0 \tag{1.3-24}$$

From Eq. (1.3-22) we obtain

$$B/A = -\delta/\kappa \tag{1.3-25}$$

The eigenvalue equation can be written in a different form:

$$\tan \kappa d = \kappa(\gamma+\delta)/(\kappa^2-\gamma\delta) \tag{1.3-26}$$

We have thus rederived the eigenvalue Eq. (1.2-16) by the precise methods of this section. The agreement of Eq. (1.3-26) with Eq. (1.2-16) justifies the heuristic method that was used in the previous section.

For future use it is convenient to express the eigenvalue equation in two alternate forms:

$$\cos \kappa d = \pm(\kappa^2-\gamma\delta)/[(\kappa^2+\gamma^2)(\kappa^2+\delta^2)]^{1/2} \tag{1.3-27}$$

and

$$\sin \kappa d = \pm\kappa(\gamma+\delta)/[(\kappa^2+\gamma^2)(\kappa^2+\delta^2)]^{1/2} \tag{1.3-28}$$

The sign of the square root must be the same in both equations. The eigenvalue equation determines the allowed values of the propagation constant β that enters Eq. (1.3-26), (1.3-27), or (1.3-28) through Eqs. (1.2-13)–(1.2-15). Equation (1.2-6) defines for each guided mode a mode angle, which is the direction of the ray or plane wave that travels inside the core region of the waveguide with refractive index n_1.

The field treatment of the guided wave problem also provides the justification for the assumption used in the preceding section that the field inside the core can be regarded as a superposition of two plane waves. Expression (1.3-17) for the E_y component of the field in the core consists of sine and cosine functions. Each of these functions can be decomposed into exponential functions of the form $\exp(\pm i\kappa x)$. If we reinstate the omitted factor (1.3-8), we see that the field in the core is formed by the superposition of plane waves of the form

$$\exp i(\omega t \mp \kappa x - \beta z) = \exp i(\omega t - \mathbf{K}\cdot\mathbf{r}) \tag{1.3-29}$$

The propagation vector $\mathbf{K}$ can be expressed as

$$\mathbf{K} = \pm\kappa\mathbf{e}_x + 0\mathbf{e}_y + \beta\mathbf{e}_z \tag{1.3-30}$$

with $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ indicating unit vectors in x, y, and z directions. The vector $\mathbf{r}$ points from the coordinate origin to the point at which the field is being considered:

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z \tag{1.3-31}$$

According to Eq. (1.2-13) we have

$$K^2 = \kappa^2 + \beta^2 = n_1^2k^2 \tag{1.3-32}$$

with K indicating the magnitude of the vector $\mathbf{K}$. This discussion makes it clear that the eigenvalue β—the propagation constant of the guided slab waveguide modes—is the z component of the propagation vector of plane waves traveling in the core. Because of condition (1.1-1) the parameter γ of Eq. (1.2-14) becomes imaginary as β becomes larger than n_2k. At the point

$$\beta = n_2k \tag{1.3-33}$$

we have

$$\gamma = 0 \tag{1.3-34}$$

We see from the field expression (1.3-18) that the field extends undiminished to infinite distances below the waveguide core, if $\gamma = 0$. When γ becomes imaginary the evanescent field in the substrate region turns into a radiation field, and the wave is no longer guided by the dielectric waveguide. We say that the wave reaches its cutoff point or simply that the wave is cutoff when Eq. (1.3-34) or (1.3-33) is reached. The cutoff condition (1.3-33) can easily be shown to be identical to the condition for the loss of total internal reflection from the dielectric interface at $x = -d$. By eliminating β from Eqs. (1.2-6) and (1.3-33) we obtain the condition for the critical angle, Eq. (1.2-3). This proves that cutoff is identical with the loss of total internal reflection. However, this statement is true only for the slab waveguide and does not necessarily hold for the round optical fiber to be discussed later.

We can obtain some rough information about the solutions of the eigenvalue equation by considering Fig. 1.3.1. The solid lines in this figure are the branches of the tangent as a function of κd. The dashed lines represent the

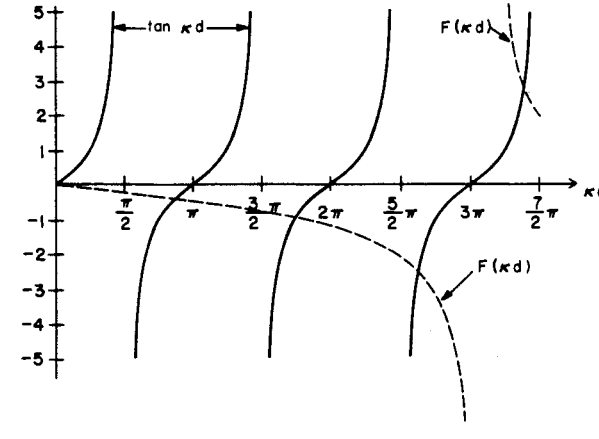

Fig. 1.3.1 *Graphical solution of the eigenvalue Eq. (1.3-26). The crossing points of the solid and dashed lines correspond to solutions.*

function $F(\kappa d)$ that represents the right-hand side of the eigenvalue Eq. (1.3-26). From Eqs. (1.2-13)–(1.2-15) we obtain

$$F(\kappa d) = \frac{\kappa d(\gamma d + \delta d)}{(\kappa d)^2 - (\gamma d)(\delta d)}$$

$$= \frac{\kappa d\{[(n_1^2 - n_2^2)(kd)^2 - (\kappa d)^2]^{1/2} + [(n_1^2 - n_3^2)(kd)^2 - (\kappa d)^2]^{1/2}\}}{(\kappa d)^2 - [(n_1^2 - n_2^2)(kd)^2 - (\kappa d)^2]^{1/2}[(n_1^2 - n_3^2)(kd)^2 - (\kappa d)^2]^{1/2}} \tag{1.3-35}$$

The figure was drawn for $(n_1^2 - n_2^2)^{1/2} kd = 11$, and $(n_1^2 - n_3^2)^{1/2} kd = 24$. The pole in the $F(\kappa d)$ curve occurs at the point where the denominator of Eq. (1.3-35) vanishes. The $F(\kappa d)$ curve ends at the point

$$(n_1^2 - n_2^2)^{1/2} kd = \kappa d \tag{1.3-36}$$

since one of the square root expressions in Eq. (1.3-35) becomes imaginary as κd exceeds the value given by Eq. (1.3-36). The κd coordinates of the crossing points of the solid and dashed curves represent solutions of the eigenvalue Eq. (1.3-26). Each solution gives one TE mode of the slab waveguide. For the conditions that were used to draw Fig. 1.3.1 we obtain four guided modes. We define a parameter that combines the difference of the squares of the refractive indices of core and medium 2 with information about the operating wavelength and the width of the core:

$$V = (n_1^2 - n_2^2)^{1/2} kd \tag{1.3-37}$$

As the value of V decreases, the endpoint of the dashed curve moves to the left, so that it crosses fewer branches of the tangent function. For decreasing values of V the number of guided modes is reduced. If V becomes small enough the endpoint of the dashed curve moves to the left of the first branch of the tangent function, so that there are no crossings of the solid and dashed lines. This means that for sufficiently narrow cores, low frequencies, or sufficiently small refractive index differences, no guided modes can exist. However, if the refractive indices of the media above and below the core are equal, $n_3 = n_2$, the endpoint of the dashed curve falls on the κd axis. In this case the dashed curve must always cross at least the first branch of the tangent curve so that at least one guided mode always exists. The symmetric slab waveguide is thus fundamentally different from the asymmetric slab in that it always supports at least one guided mode. If $n_3 = n_2$, we have $\gamma = \delta$, and Eq. (1.3-26) reduces to the simpler form

$$\tan(2\kappa d/2) = \frac{2\tan \kappa d/2}{1 - \tan^2 \kappa d/2} = \frac{2\gamma/\kappa}{1 - (\gamma/\kappa)^2} \tag{1.3-38}$$

This is a second-order equation in $\tan \kappa d/2$ with the solutions

$$\tan \kappa d/2 = \gamma/\kappa \tag{1.3-39}$$

and

$$\tan \kappa d/2 = -\kappa/\gamma \tag{1.3-40}$$

Equation (1.3-39) is the well-known eigenvalue equation for the even modes of the symmetric slab waveguide, while Eq. (1.3-40) provides the propagation constants for the odd modes [Mel]. The fact that even and odd modes result can be verified from Eq. (1.3-17) by shifting the coordinate origin to the center of the core of the symmetric slab. Descriptions of the symmetric slab waveguide usually use the symbol d for the halfwidth of the core, which explains why $d/2$ appears in Eqs. (1.3-39) and (1.3-40) instead of the customary d.

The cutoff value of V for each guided TE mode can be obtained from Eq. (1.3-26). As we pointed out earlier [see Eq. (1.3-34)] $\gamma = 0$ is the cutoff point. We have at the cutoff point of every mode the relation

$$V_c = (\kappa d)_c \tag{1.3-41}$$

We thus obtain from Eqs. (1.2-13), (1.2-15), (1.3-33), and (1.3-26)

$$V_c = \arctan[(n_2^2 - n_3^2)^{1/2}/(n_1^2 - n_2^2)^{1/2}] + \nu\pi \tag{1.3-42}$$

if the arctangent function is restricted to the range 0–$\pi/2$.† For the symmetric slab waveguide we find immediately

$$V_c = \nu\pi \tag{1.3-43}$$

or

$$(\kappa d/2)_c = \nu\pi/2 \tag{1.3-44}$$

for integer values of $\nu = 0, 1, 2, \ldots$.

It remains to relate the amplitude coefficient A of the electromagnetic field to the power carried by the mode. The power is obtained by integrating the z component of the power flow vector (Poynting vector):

$$S_z = \tfrac{1}{2} \operatorname{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot \mathbf{e}_z \tag{1.3-45}$$

over the infinite transverse cross section of the waveguide. This notation implies that S_z is a time averaged quantity. We thus obtain

$$(\beta/|\beta|)\, P = -\frac{1}{2}\int_{-\infty}^{\infty} E_y H_x^*\, dx = (\beta/2\omega\mu_0)\int_{-\infty}^{\infty} |E_y|^2\, dx \tag{1.3-46}$$

where P is a real, positive quantity. The asterisk indicates complex conjugation. For the slab with its infinite extension in y direction, P is actually the power per

† Note that we assume $n_2 \geq n_3$.

unit length (unit length in y direction). The straightforward evaluation of Eq. (1.3-46) results in

$$A^2 = 4\kappa^2\omega\mu_0 P/\{|\beta|[d+(1/\gamma)+(1/\delta)](\kappa^2+\delta^2)\} \tag{1.3-47}$$

Equations (1.3-27) and (1.3-28) were used to express the amplitude coefficient in this simple form. Throughout this book we normalize the modes so that the mode amplitudes are real quantities.

Guided TM Modes

Transverse magnetic, or TM, modes have the field components H_y, E_x, and E_z. Assuming again that the time and z dependence of the modes is given by the factor (1.3-8) we obtain from Maxwell's Eqs. (1.3-2) and (1.3-3)

$$i\beta H_y = i\omega\varepsilon_0 n^2 E_x \tag{1.3-48}$$

$$\frac{\partial H_y}{\partial x} = i\omega\varepsilon_0 n^2 E_z \tag{1.3-49}$$

$$i\beta E_x + \frac{\partial E_z}{\partial x} = i\omega\mu_0 H_y \tag{1.3-50}$$

The electric field components can now be expressed in terms of the H_y component:

$$E_x = (i/n^2\omega\varepsilon_0)\,\partial H_y/\partial z = (\beta/n^2\omega\varepsilon_0)\,H_y \tag{1.3-51}$$

and

$$E_z = (-i/n^2\omega\varepsilon_0)\,\partial H_y/\partial x \tag{1.3-52}$$

Substitution of these two equations into Eq. (1.3-50) yields the one-dimensional reduced wave equation for the H_y component:

$$(\partial^2 H_y/\partial x^2) + (n^2k^2-\beta^2)\,H_y = 0 \tag{1.3-53}$$

The boundary conditions now require that the H_y and E_z components be continuous at the dielectric interfaces $x = 0$ and $x = -d$. The solution of Eq. (1.3-53) that is continuous at the two dielectric interfaces and vanishes at $x = \pm\infty$ is given by

$$H_y = (\beta/|\beta|)\,Ce^{-\delta x}, \qquad \text{for } x \geqslant 0 \tag{1.3-54}$$

$$= (\beta/|\beta|)\,(C\cos\kappa x + D\sin\kappa x), \qquad \text{for } 0 \geqslant x \geqslant -d \tag{1.3-55}$$

$$= (\beta/|\beta|)\,(C\cos\kappa d - D\sin\kappa d)\,e^{\gamma(x+d)}, \qquad \text{for } x \leqslant -d \tag{1.3-56}$$

The factor $\beta/|\beta|$ is incorporated into the field amplitude to ensure that the transverse magnetic field changes its sign when the propagation direction is reversed [compare Eq. (3.3-11)]. The E_z component is obtained from Eq. (1.3-52):

$$E_z = (i\delta/{n_3}^2\omega\varepsilon_0)\,(\beta/|\beta|)\,Ce^{-\delta x}, \qquad \text{for } x \geqslant 0 \tag{1.3-57}$$

$$= (i\kappa/{n_1}^2\omega\varepsilon_0)\,(\beta/|\beta|)\,(C\sin\kappa x - D\cos\kappa x), \qquad \text{for } 0 \geqslant x \geqslant -d \tag{1.3-58}$$

$$= (-i\gamma/{n_2}^2\omega\varepsilon_0)\,(\beta/|\beta|)\,(C\cos\kappa d - D\sin\kappa d)\,e^{\gamma(x+d)}, \qquad \text{for } x \leqslant -d \tag{1.3-59}$$

The E_z component does not satisfy the boundary conditions at the two dielectric interfaces for arbitrary values of C and D. The requirement of continuity of E_z at $x = 0$ and $x = -d$ leads to the equation system

$$(\delta/{n_3}^2)\,C + (\kappa/{n_1}^2)\,D = 0 \tag{1.3-60a}$$

$$[(\kappa/{n_1}^2)\sin\kappa d - (\gamma/{n_2}^2)\cos\kappa d]\,C + [(\kappa/{n_1}^2)\cos\kappa d + (\gamma/{n_2}^2)\sin\kappa d]\,D = 0 \tag{1.3-60b}$$

The first equation provides a relation between the two amplitude constants

$$D/C = -({n_1}^2/{n_3}^2)\delta/\kappa \tag{1.3-61}$$

The requirement that the system determinant must vanish leads to the eigenvalue equation

$$(\delta/{n_3}^2)[(\kappa/{n_1}^2)\cos\kappa d + (\gamma/{n_2}^2)\sin\kappa d] - (\kappa/{n_1}^2)[(\kappa/{n_1}^2)\sin\kappa d - (\gamma/{n_2}^2)\cos\kappa d] = 0 \tag{1.3-62}$$

If we divide this equation by $\cos\kappa d$ and group the terms differently, we obtain again the eigenvalue Eq. (1.2-17):

$$\tan\kappa d = {n_1}^2\kappa({n_3}^2\gamma + {n_2}^2\delta)/({n_2}^2{n_3}^2\kappa^2 - {n_1}^4\gamma\delta) \tag{1.3-63}$$

We can use this equation to express the cosine and sine functions of κd in terms of the mode parameters κ, γ, and δ:

$$\cos\kappa d = \pm({n_2}^2{n_3}^2\kappa^2 - {n_1}^4\gamma\delta)/[({n_2}^4\kappa^2 + {n_1}^4\gamma^2)({n_3}^4\kappa^2 + {n_1}^4\delta^2)]^{1/2} \tag{1.3-64}$$

and

$$\sin\kappa d = \pm{n_1}^2\kappa({n_3}^2\gamma + {n_2}^2\delta)/[({n_2}^4\kappa^2 + {n_1}^4\gamma^2)({n_3}^4\kappa^2 + {n_1}^4\delta^2)]^{1/2} \tag{1.3-65}$$

Equations (1.3-63)–(1.3-65) are alternate versions of the eigenvalue equation for TM modes of the slab waveguide.

The solutions of Eq. (1.3-63) can be visualized with the help of Fig. 1.3.1. The dashed curve is slightly shifted for the TM mode case, but the principal features of the diagram are the same. For the special case of the symmetric slab we obtain from Eq. (1.3-63), with $n_2 = n_3$ for the even modes,

$$\tan(\kappa d/2) = (n_1^2/n_2^2)\gamma/\kappa \tag{1.3-66}$$

and for the odd modes

$$\tan(\kappa d/2) = -(n_2^2/n_1^2)\kappa/\gamma \tag{1.3-67}$$

At cutoff the relation (1.3-41) applies. However, the cutoff condition for TM modes is different from that of TE modes. Instead of Eq. (1.3-42) we now obtain from Eqs. (1.3-33), (1.3-34), and (1.3-63)

$$V_c = \arctan[(n_1^2/n_3^2)(n_2^2-n_3^2)^{1/2}/(n_1^2-n_2^2)^{1/2}] + v\pi \tag{1.3-68}$$

The arctangent function is again restricted to the range 0–$\pi/2$, v is either zero or an integer number, and $n_2 \geqslant n_3$ has been assumed. For the symmetric slab, cutoff conditions (1.3-43) and (1.3-44) apply also to TM modes.

Cutoff conditions (1.3-42) and (1.3-68) can be used to calculate the total number of modes that can propagate on the slab waveguide. For any given value of the parameter V defined by Eq. (1.3-37), we must require $V < V_{c,N+1}$, with $V_{c,N+1}$ indicating the cutoff value of V for mode $N+1$, which is the first mode that is no longer guided. Since $v = 0$ corresponds to the first mode, $v = N$ corresponds to mode $N+1$. For TE modes we thus obtain from Eq. (1.3-42) the inequality

$$V < N\pi + \arctan[(n_2^2-n_3^2)^{1/2}/(n_1^2-n_2^2)^{1/2}] \tag{1.3-69}$$

For the total number of TE or TM modes we have

$$N = [(1/\pi)\{V - \arctan[\eta(n_2^2-n_3^2)^{1/2}/(n_1^2-n_2^2)^{1/2}]\}]_{\text{int}} \tag{1.3-70}$$

The symbol $[\ ]_{\text{int}}$ indicates that the integer, whose value is just larger than the value of the number in brackets, must be taken. The parameter η is defined as

$$\eta = \begin{cases} 1, & \text{for TE modes} \\ n_1^2/n_3^2, & \text{for TM modes} \end{cases} \tag{1.3-71}$$

The total number of TE and TM modes is usually twice the number N given by Eq. (1.3-70) for TE or TM modes. However, it is possible that the number of TE modes is larger than the number of TM modes, because we always have $n_1/n_3 > 1$. In that case the total number of modes is the sum of the number of TE modes plus the number of TM modes that are calculated from Eq. (1.3-70).

The amplitude coefficient C can again be related to the total power that is carried by the mode. From Eq. (1.3-45) we obtain, with the help of Eq. (1.3-51),

$$\frac{\beta}{|\beta|}P = \frac{1}{2}\int_{-\infty}^{\infty} E_x H_y^*\,dx = (\beta/2\omega\varepsilon_0)\int_{-\infty}^{\infty}[1/n^2(x)]|H_y|^2\,dx \tag{1.3-72}$$

The notation $n(x)$ serves as a reminder that the refractive index is different in the three sections of the waveguide. With the field expressions (1.3-54)–(1.3-56) we obtain from Eq. (1.3-72), with the help of Eqs. (1.3-64) and (1.3-65),

$$C^2 = \frac{4\omega\varepsilon_0 P}{|\beta|} n_1^2 n_3^4 \kappa^2 \cdot \left\{(n_3^4\kappa^2 + n_1^4\delta^2)\left[d + \frac{n_1^2 n_2^2}{\gamma}\frac{\kappa^2+\gamma^2}{n_2^4\kappa^2 + n_1^4\gamma^2} + \frac{n_1^2 n_3^2}{\delta}\frac{\kappa^2+\delta^2}{n_3^4\kappa^2 + n_1^4\delta^2}\right]\right\}^{-1} \tag{1.3-73}$$

A comparison of the expressions for TE and TM modes shows that the equations describing TM modes are somewhat more cumbersome than the corresponding equations for TE modes.

Approximate Solutions of the Eigenvalue Equations

The eigenvalues κ, γ, or β (only one of the three is independent of the others) must be obtained as solutions of Eq. (1.3-26) or (1.3-63). Near cutoff and far from cutoff it is possible to give approximate solutions in closed form.

We begin by listing near cutoff approximations and assume that $n_2 > n_3$, causing γ to become very small near cutoff, while δ remains finite. Right at cutoff we have $V = V_c$, with

$$V_c = \kappa_c d = (n_1^2 - n_2^2)^{1/2} k_c d \tag{1.3-74}$$

which follows from Eqs. (1.2-13) and (1.3-37), with $\beta = n_2 k$. k_c is the free-space propagation constant for the cutoff frequency of the guided mode. The cutoff value V_c is given by Eq. (1.3-42) for the TE-type eigenvalue equations and by Eq. (1.3-68) for TM-type eigenvalue equations. In order to be able to cover both cases simultaneously, we introduce the notation

$$m_j = \begin{cases} n_j, & \text{for TM case} \\ 1, & \text{for TE case} \end{cases} \tag{1.3-75}$$

The cutoff condition thus becomes

$$V_c = v\pi + \arctan[(m_1^2/m_3^2)(n_2^2-n_3^2)^{1/2}/(n_1^2-n_2^2)^{1/2}] \tag{1.3-76}$$

The eigenvalue Eqs. (1.3-26) and (1.3-63) are combined by writing

$$\tan \kappa d = m_1^2 \kappa (m_3^2 \gamma + m_2^2 \delta)/(m_3^2 m_2^2 \kappa^2 - m_1^4 \gamma \delta) \tag{1.3-77}$$

We now assume that the waveguide is operated close to cutoff of the mode of interest and use

$$V = V_c + v \tag{1.3-78}$$

with

$$v \ll 1 \tag{1.3-79}$$

The assumption that the mode is near cutoff also allows us to use

$$\gamma d \ll V_c \tag{1.3-80}$$

From Eq. (1.2-13) we obtain approximately

$$\kappa d = V_c + v - (\gamma^2 d^2/2V_c) \tag{1.3-81}$$

and from Eq. (1.2-15) we find

$$\delta d = \delta_c d - (V_c v/\delta_c d) + (\gamma^2 d/2\delta_c) \tag{1.3-82}$$

with the abbreviation

$$\delta_c d = [(n_2^2 - n_3^2)^{1/2}/(n_1^2 - n_2^2)^{1/2}] V_c \tag{1.3-83}$$

A first-order perturbation solution of Eq. (1.3-77) yields the near cutoff approximation

$$\gamma d = \left[1 + \frac{m_1^2 m_3^2 (V_c^2 + \delta_c^2 d^2)}{\delta_c d (m_3^4 V_c^2 + m_1^4 \delta_c^2 d^2)}\right] \frac{m_2^2}{m_1^2} V_c v \tag{1.3-84}$$

This approximate solution is valid only for asymmetric waveguides, so that $\delta \gg \gamma$ near cutoff. For symmetric guides a different approximation is needed. If $n_2 = n_3$, we find for the lowest-order solution, $v = 0$, of Eq. (1.3-77),

$$\gamma d = (m_1^2/m_2^2)\{[(m_2^4/m_1^4) V^2 + 1]^{1/2} - 1\} \tag{1.3-85}$$

The cutoff value is $V_c = 0$ in this case, so that we have $V = v$. For higher-order modes we have for symmetric waveguides near cutoff

$$\gamma d = \tfrac{1}{2}(m_2^2/m_1^2) V_c v \tag{1.3-86}$$

It is apparent that Eq. (1.3-85) also assumes the form Eq. (1.3-86) for very small values of $V_c \approx V = v$, but it is advantageous to use the more complicated form Eq. (1.3-85) for the lowest-order mode since it gives more accurate results.

Far from cutoff we use the fact that

$$\kappa \ll \gamma, \quad \text{and} \quad \kappa \ll \delta \tag{1.3-87}$$

We write

$$\kappa d = \kappa_\infty d/(1+\varepsilon) \tag{1.3-88}$$

with (see Fig. 1.3.1)

$$\kappa_\infty d = (\nu + 1)\pi \tag{1.3-89}$$

and with the assumption

$$\varepsilon \ll 1 \tag{1.3-90}$$

From Eq. (1.3-77) we find the approximation for ε:

$$\varepsilon = (m_3^2 \gamma_\infty + m_2^2 \delta_\infty)/m_1^2 \gamma_\infty \delta_\infty d \tag{1.3-91}$$

The notation γ_∞ and δ_∞ is used to indicate that we use $\kappa = \kappa_\infty$ of Eq. (1.3-89) to obtain the values of γ and δ from Eqs. (1.2-14) and (1.2-15).

This approximate solution for κd is remarkably accurate. We could, of course, have written $1-\varepsilon$ instead of $(1+\varepsilon)^{-1}$, but comparison with the exact solutions of the eigenvalue equations shows that better results are obtained with the form Eq. (1.3-88). We can go one step further and use a first approximation for κd obtained from Eq. (1.3-88) to calculate improved values for γ and δ from Eqs. (1.2-14) and (1.2-15) and use these to calculate improved values of ε from Eq. (1.3-91). The far from cutoff approximations are compared to the exact solutions in Sect. 1.7 describing the rectangular dielectric waveguide.

The propagation constant β is obtained either from

$$\beta = (n_1^2 k^2 - \kappa^2)^{1/2} \tag{1.3-92}$$

with the help of Eq. (1.3-88) or, for the near cutoff approximation, from Eqs. (1.3-84)–(1.3-86) and

$$\beta = (n_2^2 k^2 + \gamma^2)^{1/2} \tag{1.3-93}$$

1.4 Radiation Modes of the Asymmetric Slab Waveguide

A slab waveguide can support guided modes if expression (1.3-70) for the number of TE or TM modes is larger than zero. However, the number of guided modes is always finite so that there must be other solutions of Maxwell's equations that satisfy the boundary conditions in order to provide a complete set of orthogonal modes. Such additional modes do indeed exist. It is helpful to resort to a physical argument to illustrate the nature of the various modes of the waveguide.

In Sect. 1.2 we discussed the mechanism of mode guidance on the basis of geometrical optics. There it was assumed that a ray or plane wave already exists inside the waveguide core. Once such a ray was postulated it could be

shown that it would travel in the waveguide without loss of power, provided that losses in the dielectric material were ignored. The source of the guided mode field must be presumed to be inside the waveguide core and, if the waveguide is infinitely long, the source must be located at minus infinity. The fact that a source inside the core does not necessarily excite only guided modes shall not concern us for the moment.

Let us now assume that we place a source of radiation outside the waveguide core. It is easy to see that such a source can not contribute to guided modes if the waveguide structure is perfect. The waves emitted by the source outside the core reflect and refract at the core boundary, but none of its energy is trapped and travels inside the core as a guided wave. It would require imaginary angles in order to inject a ray from the outside into the core in such a way that it travels inside with less than the critical angle for total internal reflection. According to Snell's law, Eq. (1.2-2), the ray angle inside the core (whose refractive index is larger than that of the surrounding medium) is larger than the angle on the outside. A ray that can be refracted into the core must thus always exceed the critical angle. An imaginary ray angle corresponds to an evanescent field. Such fields can tunnel into the core from the outside but are themselves evanescent field tails resulting from total internal reflection and cannot result from the radiation field of a source outside the waveguide core.

If we visualize a plane wave impinging on the waveguide core from an infinite distance we know that a portion of this wave is reflected at the core boundary, while the remaining energy penetrates through the core and emerges on the other side as a plane wave. The direction of these plane waves can be obtained by applying Snell's law repeatedly. A plane wave impinging on the core from above thus results in a reflected wave above the core and a transmitted wave below the core. The total field in the region above the core is thus a standing wave. The resulting radiation field must, of course, be a solution of Maxwell's equations and it must also satisfy the boundary conditions. In addition, the t and z dependence of the field can again be described by factor (1.3-8). The radiation field thus qualifies in all respects as a mode, except that it is not confined to the waveguide core but reaches undiminished to infinite distances in x direction normal to the core. We call modes of this type radiation modes. Their propagation constants β are not constrained to a discrete set of values, since they are related to the angle of the incident plane wave which can be chosen arbitrarily. The values of the propagation constant thus form a continuum, so that we also speak of the radiation modes as modes of the continuum.

The radiation modes are necessary to describe radiation phenomena in the region around the waveguide core. We shall see in later chapters that waveguide imperfections cause some of the guided mode power to radiate away

into the space outside the core. However, radiation that originates inside the core results in traveling wave fields in the outside regions, while we have seen that the radiation modes must be standing waves, at least on one side of the waveguide core. This has caused considerable confusion and makes it difficult to understand the mechanism by which the standing wave radiation modes (standing waves in transverse direction, the modes travel along the z direction) can contribute to strictly outgoing radiation. Let us first reiterate that there are no normalizable solutions of Maxwell's equations satisfying the boundary conditions at the core interface that form only traveling waves outside the core regions. (Leaky waves do not form standing wave patterns but grow exponentially in x direction.) This fact can be shown mathematically, but it is obvious from our physical argument since an incident traveling wave is always partially reflected at the core boundary resulting in a standing wave. The resolution of the seeming paradox is obtained when we consider that the radiation modes form a continuum. It is impossible to excite one single radiation mode with a source inside the core of a waveguide. In fact, any mechanism that excites a continuum mode always simultaneously excites infinitely many continuum modes in its vicinity (vicinity in the sense of β space). Only an infinitely extended source at infinity could excite a pure radiation mode. However, such a source is physically impossible. If radiation is excited by an imperfection of the waveguide it always excites infinitely many radiation modes, which superimpose themselves in such a way that the incoming parts of the standing wave are eliminated by destructive interference. It is not easy to show this rigorously, but the approximate solution of the resulting integrals by the method of stationary phase shows clearly that only the outgoing waves contribute in proper phase, while the incoming wave components of the radiation modes fail to interfere constructively.

After these introductory remarks we proceed to derive the mathematical expressions of the radiation mode fields from Maxwell's equations.

TE Radiation Modes

The analysis proceeds in close analogy to the derivation of the guided modes. The number of nonvanishing field components is the same for both types of modes. The H_x and H_z field components can again be obtained from Eqs. (1.3-13) and (1.3-14) in terms of the E_y component. The E_y component is obtained from the reduced wave Eq. (1.3-15).

We can convince ourselves easily that asymmetric slab waveguides have two types of radiation modes. As always, we assume that the refractive index n_3 of the region above the core is smaller than the index n_2 of the region below the core. A wave impinging on the core from below can thus suffer total internal reflection at the interface of regions 1 and 3. In this case we

obtain an evanescent field in region 3 and a standing wave in the core and in region 2. The range of β values that belongs to this type of radiation modes follows directly from Snell's law (1.2-2). The smallest angle of the incident ray is $\theta_2 = 0$ on the outside but becomes θ_1 inside the core, so that we have

$$\beta = n_1 k \cos\theta_1 = n_2 k \tag{1.4-1}$$

The largest β value of the radiation modes thus coincides with the cutoff value (1.3-33) of the guided modes. The smallest β value which still results in total internal reflection at the upper core boundary follows again from Snell's law. This time we must require that the angle θ_3 of the emerging ray in medium 3 vanishes, so that we have

$$\beta = n_1 k \cos\theta_1' = n_3 k \tag{1.4-2}$$

The range of β values for radiation modes, which have exponentially decaying fields in region 3, is thus

$$n_3 k \leqslant |\beta| < n_2 k \tag{1.4-3}$$

This type of radiation mode with exponentially decaying fields on one side of the core is peculiar to the asymmetric slab waveguide. We see from Eq. (1.4-3) that its range shrinks to zero when $n_2 = n_3$, that is, when the waveguide is of the symmetric type. Radiation modes with β values in the range (1.4-3) are responsible for radiation phenomena with power escaping only into the substrate (that is region 2) but not into the space above the core. The electric field component E_y of radiation modes in the range (1.4-3) is described by the following expressions:

$$E_y = A_r e^{i\Delta x}, \qquad \text{for} \quad x \geqslant 0 \tag{1.4-4}$$

$$= A_r \cos\sigma x + B_r \sin\sigma x, \qquad \text{for} \quad 0 \geqslant x \geqslant -d \tag{1.4-5}$$

$$= (A_r \cos\sigma d - B_r \sin\sigma d)\cos\rho(x+d) + \bar{C}_r \sin\rho(x+d), \qquad \text{for} \quad x \leqslant -d \tag{1.4-6}$$

The constants appearing in these equations are adjusted to assure continuity of the E_y component at $x = 0$ and $x = -d$. The parameters Δ, σ, and ρ are defined by the equations

$$\Delta = (n_3^2 k^2 - \beta^2)^{1/2} \tag{1.4-7}$$

$$\sigma = (n_1^2 k^2 - \beta^2)^{1/2} \tag{1.4-8}$$

and

$$\rho = (n_2^2 k^2 - \beta^2)^{1/2} \tag{1.4-9}$$

Note that Δ is positive imaginary in the range (1.4-3). The E_y component thus decays exponentially in region 3 for $x > 0$. This notation was chosen so that we can keep the same symbol Δ also for the modes outside the range (1.4-3). The field in space 2, the substrate region, is a standing wave in accordance with our physical discussion of the origin of radiation modes. The three amplitude coefficients are necessary to keep the field expressions general. The H_z component is obtained from the E_y component with the help of Eq. (1.3-14):

$$H_z = (-\Delta/\omega\mu_0) A_r e^{i\Delta x}, \qquad \text{for} \quad x \geqslant 0 \tag{1.4-10}$$

$$= (-i\sigma/\omega\mu_0)(A_r \sin\sigma x - B_r \cos\sigma x), \qquad \text{for} \quad 0 \geqslant x \geqslant -d \tag{1.4-11}$$

$$= (-i\rho/\omega\mu_0)[(A_r \cos\sigma d - B_r \sin\sigma d)\sin\rho(x+d) - \bar{C}_r \cos\rho(x+d)], \qquad \text{for} \quad x \leqslant -d \tag{1.4-12}$$

The E_y component already satisfies the boundary conditions at $x = 0$ and $x = -d$. In order to force the H_z component to satisfy the boundary condition, requiring continuity at the two interfaces, we must satisfy the following two equations:

$$-i\sigma B_r = \Delta A_r \tag{1.4-13}$$

$$\sigma \cos\sigma d\, B_r - \rho \bar{C}_r = -\sigma \sin\sigma d\, A_r \tag{1.4-14}$$

When we determined the guided modes of the asymmetric slab waveguide we found that the boundary conditions led to a determinantal condition which furnished the eigenvalue equation from which the allowed values of the propagation constant β could be determined. Equations (1.4-13) and (1.4-14) contain three coefficients. We may regard one of them, A_r for example, as given and then consider the equation system as a set of two inhomogeneous equations from which B_r and $\bar{C}_r$ can be determined. However, the system determinant must now be nonvanishing so that no eigenvalue equation results. The values of the propagation constant β thus remain arbitrary and form a continuum in the range (1.4-3). The two amplitude coefficients can be expressed in terms of A_r:

$$B_r = (i\Delta/\sigma) A_r \tag{1.4-15}$$

$$\bar{C}_r = [(\sigma/\rho)\sin\sigma d + (i\Delta/\rho)\cos\sigma d] A_r \tag{1.4-16}$$

Radiation modes cannot be normalized with respect to a finite amount of power. If we calculate the integral of the power expression (1.3-46) for one

radiation mode as given by Eqs. (1.4-4)–(1.4-6), we find that the integral diverges. However, the normalization of continuum modes can be accomplished with the help of the Dirac delta function $\delta(x)$. Instead of Eq. (1.3-46) we require for radiation modes

$$\frac{1}{2}\int_{-\infty}^{\infty} [\hat{\mathbf{E}}(\rho)\times\hat{\mathbf{H}}^*(\rho')]\cdot\mathbf{e}_z\,dx = s_\rho(\beta^*/|\beta|)\,P\,\delta(\rho-\rho') \qquad (1.4\text{-}17)$$

where P is always real and positive.

This equation has several new features compared to Eq. (1.3-46). The two field expressions under the integral sign belong to different radiation modes. The delta function on the right-hand side states that the integral vanishes if the two modes are different, but that it becomes infinitely large if both modes are identical. Several features of the expression were introduced to allow for the fact that the propagation constant β can become imaginary, as we shall see later. The caret on top of the field quantities states that the propagation factor (1.3-7) must not be included in the field expressions. It has been our practice to omit this factor from the equations for simplicity of notation. However, we now require specifically that this factor be absent from the field expressions. For real values of β it would not make any difference whether we include Eq. (1.3-7) in the field expressions, since the complex conjugation causes this factor to cancel out for $\rho = \rho'$, and for $\rho \neq \rho'$ the integral vanishes. But for imaginary β values the term (1.3-7) would not cancel, so that the orthogonality expression would become a function of z.

The ratio $\beta^*/|\beta|$ causes the right-hand side of the equation to become negative for waves traveling in negative z direction, so that this factor assures us that P is always positive. For imaginary β values expression (1.4-17) becomes imaginary. Since the real part of the expression on the left-hand side expresses the average power flow, imaginary values of β cause no power to flow along the z axis. Finally we must explain the factor s_ρ. For real values of β we always have $s_\rho = 1$. However, for imaginary β we may have to require $s_\rho = -1$ in order to keep P positive. For TE modes we always have $s_\rho = 1$ for all possible values of β. For TE modes Eq. (1.4-17) can be written as

$$(\beta^*/2\omega\mu_0)\int_{-\infty}^{\infty} \hat{E}_y(\rho)\,\hat{E}_y^*(\rho')\,dx = (\beta^*/|\beta|)\,P\,\delta(\rho-\rho') \qquad (1.4\text{-}18)$$

The arguments ρ and ρ' label two different radiation modes. Equations (1.4-17) and (1.4-18) establish not only the normalization of the radiation modes but also their orthogonality. It can be shown by direct calculation that relation (1.4-18) is indeed true. This relation is used to express the amplitude coefficient A_r of the radiation mode in terms of the factor P appearing in the normalization and orthogonality condition (1.4-18). The actual calculation requires some care, since it is necessary to recognize the delta function in the expressions that

result from substitution of Eqs. (1.4-4)–(1.4-6) into Eq. (1.4-18). The details of such a calculation were shown in [Me1, pp. 316–317]. The calculation results in the following expression for the amplitude coefficient A_r:

$$A_r^2 = \frac{4\omega\mu_0\rho^2\sigma^2 P}{\pi|\beta|[\rho^2(\sigma\cos\sigma d - i\Delta\sin\sigma d)^2 + \sigma^2(\sigma\sin\sigma d + i\Delta\cos\sigma d)^2]} \qquad (1.4\text{-}19)$$

Next we derive the radiation modes for β values in the range

$$-n_3 k \leqslant \beta \leqslant n_3 k \qquad (1.4\text{-}20)$$

These modes correspond to plane waves impinging on the core at such angles that no total internal reflection results at the upper core interface. However, instead of using only a single plane wave incident from above or below, we assume that two sources send waves toward the core, one from above and the other from below. The reason for this choice is not so obvious in the present case of the asymmetric slab waveguide. However, this procedure results in even and odd radiation modes in the symmetric case provided the waves are properly phased. We adjust our radiation modes such that even and odd modes result in the limit $n_2 = n_3$. The following field expressions satisfy the wave Eq. (1.3-15) and the boundary conditions:

$$E_y = C_r[\cos\Delta x + (\sigma/\Delta)F_i\sin\Delta x], \qquad \text{for } x \geqslant 0 \qquad (1.4\text{-}21)$$

$$= C_r(\cos\sigma x + F_i\sin\sigma x), \qquad \text{for } 0 \geqslant x \geqslant -d \qquad (1.4\text{-}22)$$

$$= C_r[(\cos\sigma d - F_i\sin\sigma d)\cos\rho(x+d) + (\sigma/\rho)(\sin\sigma d + F_i\cos\sigma d)\sin\rho(x+d)], \qquad \text{for } x \leqslant -d \qquad (1.4\text{-}23)$$

These field expressions have already been adjusted so that the H_z component, which follows from Eq. (1.3-14), is also continuous at the two interfaces. For simplicity, the detailed expressions for H_z are not given. The parameters Δ, σ, and ρ are defined by Eqs. (1.4-7)–(1.4-9). Note however, that in the range (1.4-20) Δ is a real parameter.

The field expressions for the radiation modes, Eqs. (1.4-21)–(1.4-23), contain two undetermined constants, C_r and F_i. One of them, C_r, can again be related to the parameter P of the normalizing expression (1.4-18). However, the constant F_i remains completely arbitrary. We are free to choose F_i according to our own convenience. If we use a certain value F_1, we obtain one set of radiation modes. A second choice F_2 results in another set of radiation modes. We thus see that we obviously have obtained two independent types of radiation modes. The freedom of choice of F_i is related to the arbitrary phases of the two plane waves that generate the radiation modes.

For future applications it is most important to choose these two types of modes to be mutually orthogonal. However, even this requirement does not completely determine both F_1 and F_2. In the case of the symmetric slab waveguide this problem was avoided by choosing even and odd radiation modes from the start. In this case no undetermined parameters occur in the equations. We find it convenient to choose F_1 and F_2 in our present situation in such a way that even and odd radiation modes result in the limit $n_2 = n_3$. This choice is not dictated by necessity but only by convenience.

The requirement that even and odd radiation modes result in the limit of a symmetric slab yields the following explicity expressions for F_1 and F_2:

$$F_{1,2} = [(\sigma^2-\rho^2)\sin 2\sigma d]^{-1}\{(\sigma^2-\rho^2)\cos 2\sigma d + (\rho/\Delta)(\sigma^2-\Delta^2) \pm [(\sigma^2-\rho^2)^2 + 2(\rho/\Delta)(\sigma^2-\rho^2)(\sigma^2-\Delta^2)\cos 2\sigma d + (\rho^2/\Delta^2)(\sigma^2-\Delta^2)^2]^{1/2}\} \tag{1.4-24}$$

Both signs of the square root are used to determine F_1 or F_2. The plus sign belongs to the odd modes, while the minus sign belongs to the even modes, in the limit $n_2 = n_3$. The two types of radiation modes are obtained by using either F_1 or F_2 in the field expressions (1.4-21)–(1.4-23). It can be checked that the following relation is valid:

$$F_1 F_2 = -1 \tag{1.4-25}$$

The amplitude coefficient C_r of the radiation modes can again be related to the factor P by means of expression (1.4-18):

$$C_r^2 = \frac{4\omega\mu_0 P}{\pi|\beta|} \cdot \left[(\cos\sigma d - F_i\sin\sigma d)^2 + \frac{\sigma^2}{\rho^2}(\sin\sigma d + F_i\cos\sigma d)^2 + \left(1+\frac{\sigma^2}{\Delta^2}F_i^2\right)\frac{\Delta}{\rho}\right]^{-1} \tag{1.4-26}$$

Actually C_r also needs the label $i = 1$ or $i = 2$, since both values of i are used to label F_i in its denominator. We omit these additional labels in order to keep the notation simpler.

The parameter ρ is used to label the radiation modes [see Eq. (1.4-18)]. It is allowed to assume all values from 0 to ∞. If

$$0 \leqslant \rho \leqslant (n_2^2 - n_3^2)^{1/2}k \tag{1.4-27}$$

β covers the range of values given by Eq. (1.4-3), so that this range of the parameter ρ belongs to the radiation modes (1.4-4)–(1.4-6). As ρ covers the range

$$(n_2^2 - n_3^2)^{1/2}k \leqslant \rho \leqslant n_2 k \tag{1.4-28}$$

the corresponding β values lie in the range (1.4-20) belonging to the modes (1.4-21)–(1.4-23). However, ρ is also allowed to fall in the range

$$n_2 k \leqslant \rho < \infty \tag{1.4-29}$$

It is apparent from Eq. (1.4-9) that the corresponding β values are imaginary. The radiation modes that correspond to imaginary β values are also described by Eqs. (1.4-21)–(1.4-23). These modes decay exponentially along the z axis and are necessary to describe the fine structure of the field in the immediate vicinity of a waveguide imperfection. Radiation modes of this kind cannot be generated by a plane wave source at infinity. We have thus found an extension of the simple intuitive range of radiation modes whose generation could easily be visualized by physical arguments. Radiation modes with evanescent fields along the z direction are not very important for most practical applications. However, they are necessary to form a complete set of orthogonal modes which is capable of expressing any field distribution satisfying Maxwell's equations. Evanescent waves of this type are familiar from the theory of hollow metallic waveguides, where they are associated with waves beyond cutoff. Cutoff has a different meaning in dielectric waveguides and is not associated with evanescent waves but with fields that lose power continuously by radiation. Such leaky waves will be discussed in Sect. 1.5. However, we can convince ourselves that the evanescent waves of our dielectric waveguide do have a close relationship to the cutoff waves in metallic waveguides. Let us assume that we enclose the slab waveguide with plane metallic plates on both sides. The two metal surfaces are assumed to be perfect conductors. The new structure is a metallic waveguide with a dielectric insert, all the modes of which belong to a discrete spectrum of β values. A finite number of them can propagate, while an infinite number is beyond cutoff in the usual waveguide sense. These cutoff modes decay exponentially in z direction. As we allow the metal surfaces to move further away from the core of the dielectric slab, but keep the region between core and metal plates filled with the two media with refractive indices n_2 and n_3, we find that the modes of the structure assume two different features. We again obtain the usual guided modes of the dielectric slab waveguide which are unaffected by the presence of the metal surfaces, provided these are sufficiently far away so that the exponentially decaying fields (in transverse direction) reach them with practically zero intensity. In addition to these surface modes of the dielectric slab, we find modes of the metallic waveguide that fill the entire volume between the two reflectors. The β values of these latter modes become closer spaced as the metal surfaces recede more from the core region and become the continuum of radiation modes in the limit of infinitely far reflectors. The cutoff modes of the metallic waveguide remain cutoff even if the reflectors are infinitely far spaced. (This is true only if we move the metal plates in discrete jumps so that they are placed at

successive zeros of the mode we wish to discuss.) The modes are cutoff in the sense of metallic waveguides if their transverse nodes follow each other at distances that are spaced closer than half the plane wave wavelength in the medium filling the guide. The nature of the evanescent radiation modes is thus identical to the cutoff modes of metallic waveguides. They are evanescent because of their rapid transverse variation.

TM Radiation Modes

The TM radiation modes have the same features that we encountered in discussing TE radiation modes. With restriction (1.3-4) we have only three nonvanishing field components H_y, E_x, and E_z. The H_y component is obtained from the reduced wave Eq. (1.3-53), and the components E_x and E_z can be calculated from Eqs. (1.3-51) and (1.3-52). We list only the H_y component and find radiation modes in the range (1.4-3) [for a definition of Δ, σ, and ρ see Eqs. (1.4-7)–(1.4-9)]:

$$H_y = (\beta/|\beta|)\, D_r e^{i\Delta x}, \qquad \text{for } x \geqslant 0 \tag{1.4-30}$$

$$= (\beta/|\beta|)(D_r \cos\sigma x + G_r \sin\sigma x), \qquad \text{for } 0 \geqslant x \geqslant -d \tag{1.4-31}$$

$$= (\beta/|\beta|)[(D_r \cos\sigma d - G_r \sin\sigma d)\cos\rho(x+d) + K_r \sin\rho(x+d)], \qquad \text{for } x \leqslant -d \tag{1.4-32}$$

The H_y component has been adjusted to satisfy the boundary conditions at $x = 0$ and $x = -d$. The requirement that E_z also satisfy the boundary conditions leads to the determination of the coefficients G_r and K_r:

$$G_r = (n_1^2/n_3^2)(i\Delta/\sigma) D_r \tag{1.4-33}$$

and

$$K_r = [(n_2^2/n_1^2)(\sigma/\rho)\sin\sigma d + (n_2^2/n_3^2)(i\Delta/\rho)\cos\sigma d] D_r \tag{1.4-34}$$

From Eq. (1.4-17) we obtain for TM modes, with the help of Eq. (1.3-51),

$$(\beta/2\omega\varepsilon_0)\int_{-\infty}^{\infty} [1/n^2(x)] H_y(\rho)\, H_y^*(\rho')\, dx = s_\rho(\beta^*/|\beta|)\, P\delta(\rho-\rho') \tag{1.4-35}$$

The function $n(x)$ assumes the constant values n_1, n_2, and n_3 in the three regions of the slab. The normalization condition and orthogonality relation (1.4-35) allows us to express D_r in terms of P:

$$D_r^2 = (4\omega\varepsilon_0 n_1^4 n_2^2 n_3^4 \rho^2\sigma^2 P/\pi|\beta|)(\beta^*/\beta)s_\rho \cdot [n_1^4\rho^2(n_3^2\sigma\cos\sigma d - n_1^2 i\Delta\sin\sigma d)^2 + n_2^4\sigma^2(n_3^2\sigma\sin\sigma d + n_1^2 i\Delta\cos\sigma d)^2]^{-1} \tag{1.4-36}$$

(The factor s_ρ must be chosen $+1$ or -1 to keep D_r^2 positive.) The TM radiation modes of this first type decay exponentially in region 3, for $x > 0$, since $i\Delta$ is a real negative quantity for β values in the range (1.4-3).

The second type of radiation modes consists of standing waves above and below the core region. In complete analogy to the TE modes of this type, we now have

$$H_y = (\beta/|\beta|)\, S_r[\cos\Delta x + (n_3^2/n_1^2)(\sigma/\Delta)\, R_i \sin\Delta x] \qquad \text{for } x \geqslant 0 \tag{1.4-37}$$

$$= (\beta/|\beta|)\, S_r(\cos\sigma x + R_i \sin\sigma x), \qquad \text{for } 0 \geqslant x \geqslant -d \tag{1.4-38}$$

$$= (\beta/|\beta|)\, S_r[(\cos\sigma d - R_i\sin\sigma d)\cos\rho(x+d) + (n_2^2/n_1^2)(\sigma/\rho)(\sin\sigma d + R_i\cos\sigma d)\sin\rho(x+d)], \qquad \text{for } x \leqslant -d \tag{1.4-39}$$

The parameters Δ, σ, and ρ are defined by Eqs. (1.4-7)–(1.4-9), and Δ is a real quantity in the range (1.4-28) and (1.4-29) which belongs to radiation modes of this type.

We choose R_i so that modes 1 and 2 are orthogonal and even and odd radiation modes result in the limit $n_2 = n_3$ of the symmetric slab:

$$R_{1,2} = [(n_2^4\sigma^2 - n_1^4\rho^2)\sin 2\sigma d]^{-1} \cdot \{(n_2^4\sigma^2 - n_1^4\rho^2)\cos 2\sigma d + (n_2^2/n_3^2)(\rho/\Delta)(n_3^4\sigma^2 - n_1^4\Delta^2) \pm [(n_2^4\sigma^2 - n_1^4\rho^2)^2 + (n_2^4/n_3^4)(\rho^2/\Delta^2)(n_3^4\sigma^2 - n_1^4\Delta^2)^2 + 2(n_2^2/n_3^2)(\rho/\Delta)(n_2^4\sigma^2 - n_1^4\rho^2)(n_3^4\sigma^2 - n_1^4\Delta^2)\cos 2\sigma d]^{1/2}\} \tag{1.4-40}$$

R_1 is obtained for the positive sign of the square root, while R_2 belongs to the negative sign. The plus sign leads to odd modes, and the minus sign to even modes in the limit of a symmetric slab, $n_2 = n_3$. It must be remembered that even and odd field distributions are referred to the center of the core. In order to see that even and odd modes do indeed result from R_2 and R_1, it is necessary to transform Eqs. (1.4-37)–(1.4-39) to a coordinate system that is centered in the middle of the core.

The amplitude coefficients of the modes are obtained by substituting the field expressions (1.4-37)–(1.4-39) into Eq. (1.4-35). The tedious calculation results in

$$S_r^2 = (4\omega\varepsilon_0 P/\pi|\beta|)(\beta^*/\beta)s_\rho \cdot \left[\frac{1}{n_2^2}(\cos\sigma d - R_i\sin\sigma d)^2 + \frac{n_2^2}{n_1^4}\frac{\sigma^2}{\rho^2}(\sin\sigma d + R_i\cos\sigma d)^2 + \left(\frac{1}{n_3^2} + \frac{n_3^2}{n_1^4}\frac{\sigma^2}{\Delta^2}R_i^2\right)\frac{\Delta}{\rho}\right]^{-1} \tag{1.4-41}$$

We must choose $s_\rho = +1$ or -1 to keep S_r^2 positive. The radiation modes will be needed in later sections to calculate radiation losses caused by waveguide irregularities.

Mode Orthogonality

All waveguide modes are mutually orthogonal to each other. Orthogonality is defined in the sense of Eq. (1.4-17). If we use any guided or radiation mode for the field labeled $E(\rho)$ and any other guided or radiation mode for $H(\rho')$, we find that the integral over the z component of the vector product of the two mode fields vanishes. The labels ρ and ρ' are here used to distinguish between two different modes. These labels may indicate either guided modes, in which case they are discrete quantities, or radiation modes, in which case they form a continuum. Only if $E(\rho)$ and $H(\rho')$ belong to the same mode is the result of the integration different from zero. For guided modes we can normalize the field to make the right-hand side equal to the power P carried by the mode. We always normalize our normal modes so that they carry the same amount of power, for example 1 W. The function $\delta(\rho-\rho')$ on the right-hand side of Eq. (1.4-17) indicates a Dirac delta function when both ρ and ρ' label continuum modes. For guided modes we must interpret this symbol as the Kronecker delta, which is unity when both indices are equal and zero otherwise. The orthogonality relation (1.4-17) has the physical meaning that the power carried by the waveguide field is simply the sum of the powers carried by all the modes. If we consider the superposition of several modes and form the power of the mode field by calculating the integral of the Poynting vector over the infinite cross section, we can use Eq. (1.4-17) to convert the expression for the total power to the sum of the powers carried by each mode. This is not a trivial statement. In fact, mode orthogonality with respect to average power as expressed by Eq. (1.4-17) applies only for lossless dielectric waveguides. If the refractive indices n_1, n_2, and n_3 are complex quantities, relation (1.4-17) does not hold. It will still hold to a good approximation for only slightly lossy waveguides, but it is no longer strictly true. However, it can be shown that another orthogonality relation applies, which follows from Eq. (1.4-17) simply by omitting the complex conjugation from the H field. This mode orthogonality is more general than the power orthogonality implied by Eq. (1.4-17). Without the complex conjugation the orthogonality relation holds for lossy as well as lossless waveguides, but the simple interpretation of power orthogonality is lost in this case. The quantity $\mathbf{E}\times\mathbf{H}$ without the complex conjugate for $\mathbf{H}$ does not have a physical meaning. This is the reason we prefer the form (1.4-17) of the orthogonality relation. The validity of Eq. (1.4-17) has been proven in many places (see, for example, Sect. 8.5 of [Mel]). Both types of orthogonality relations can, of course, be verified by direct calculation using the field expressions for the modes.

1.5 Leaky Waves

The guided and radiation modes, whose field expressions we derived in the last two sections, form a complete orthogonal set of modes. Any field distribution that obeys the restriction (1.3-4) can be expressed by series expansion into these modes. However, there are other solutions of the eigenvalue Eqs. (1.3-26) and (1.3-63) that are not part of the complete set of modes [Sol]. These additional solutions do not belong to proper field expressions, since the fields associated with them diverge at $x = \pm\infty$. The integrals in Eqs. (1.3-46) and (1.3-72) diverge and, since these modes have a discrete spectrum, they cannot even be normalized with the help of the Dirac delta function. However, if we do not try to compute the total power carried by these modes and if we ignore the normalization problem, we can interpret them as representing guided modes beyond the cutoff point.

It is helpful to use physical intuition to understand leaky waves, as these additional solutions of the eigenvalue problem are called. Using the geometrical optics argument developed in Sect. 1.2 we can obtain an eigenvalue equation for leaky waves. Since we are now considering waves that are beyond cutoff from the point of view of the usual guided modes, we must restrict ourselves to propagation constants in the range

$$-n_2 k < \beta < n_2 k \tag{1.5-1}$$

This range of β values belongs to the domain of radiation modes. The leaky waves to be considered here have thus the same propagation constants as the radiation modes, except that they form a discrete set instead of a continuum. In addition, leaky waves are always lossy so that their propagation constants have complex values. We are here about to derive an approximation to the real part of the complex propagation constants of leaky waves.

In the β range (1.5-1) both square roots appearing in Eq. (1.2-5) are real, so that there is no phase shift on reflection from the dielectric interface with region 2. With $\phi_2 = 0$ we obtain from Eq. (1.2-11) with the help of Eqs. (1.2-9), (1.2-10), (1.2-13), and (1.3-75) the eigenvalue condition

$$\kappa d = \begin{cases} N\pi + \arctan[(m_1^2/m_3^2)(\delta/\kappa)], & \text{for } \delta^2 > 0 \\ N\pi, & \text{for } \delta^2 < 0 \end{cases} \tag{1.5-2}$$

The real part of the propagation constant is then obtained from Eqs. (1.2-13) and (1.5-2):

$$\beta = (n_1^2 k^2 - \kappa^2)^{1/2} \tag{1.5-3}$$

This result holds for TE (with $m_i = 1$) and for TM modes (with $m_i = n_i$).

We obtain the imaginary part of the propagation constant by considering the reflection coefficient of plane waves that superimpose themselves on each other to form the leaky wave in the core. If we use the notation

$$\kappa = (n_1^2 k^2 - \beta^2)^{1/2} \tag{1.5-4}$$

$$\rho = (n_2^2 k^2 - \beta^2)^{1/2} \tag{1.5-5}$$

and

$$\Delta = (n_3^2 k^2 - \beta^2)^{1/2} \tag{1.5-6}$$

we obtain the power reflection coefficient from (1.2-5):

$$R_2 = |r|^2 = [(\kappa - \rho)/(\kappa + \rho)]^2 \tag{1.5-7}$$

for reflection from medium 2 and†

$$R_3 = |(\kappa - \Delta)/(\kappa + \Delta)|^2 \tag{1.5-8}$$

for reflection from medium 3. Concentrating on one interface for the moment, we obtain the power ΔP, that is lost per unit length, from the following equation:

$$\Delta P_2 = S(1 - R_2) \sin\theta \tag{1.5-9}$$

where S is the magnitude of the Poynting vector of the plane wave and indicates the amount of power that is flowing through the unit area in the direction of plane wave propagation. $1 - R$ is the ratio of the power flowing normal to the surface out of region 1 divided by the amount of power that is carried normal to that surface. Multiplication of $1 - R$ by the x component, $S_x = S \sin\theta$, of the Poynting vector yields the power that is lost per unit length of the waveguide (see Fig. 1.5.1). Similarly we find the power loss per unit length on the other interface

$$\Delta P_3 = S(1 - R_3) \sin\theta \tag{1.5-10}$$

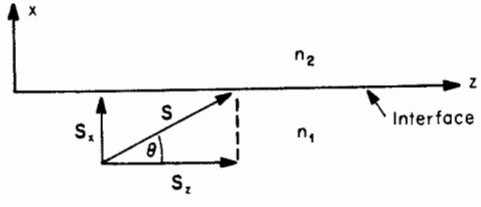

Fig. 1.5.1 *Illustration of the decomposition of the energy flow vector into components parallel and perpendicular to the dielectric interface.*

† Δ may be real or imaginary, but ρ is real in the interval (1.5-1).

The power loss coefficient of the leaky wave is defined as the amount of power that is lost per unit length divided by the power carried in the guide. The power carried by the plane wave inside the waveguide is

$$P = S_z d = Sd \cos\theta \tag{1.5-11}$$

The power loss coefficient 2α is thus

$$2\alpha = (\Delta P_2 + \Delta P_3)/2P \tag{1.5-12}$$

The factor 2 in the denominator is needed to adjust for the fact that the combined length over which the power $\Delta P_2 + \Delta P_3$ is dissipated is twice the unit length. We define α as the amplitude loss coefficient, so that 2α is the power loss. From Eqs. (1.5-7)–(1.5-12) we finally obtain, with the help of Eqs. (1.2-6) and (1.2-13), the power loss coefficient for leaky TE waves:

$$2\alpha = \frac{2}{d}\left[\frac{\rho}{(\kappa+\rho)^2} + \frac{\Delta + \Delta^*}{2|\kappa + \Delta|^2}\right]\frac{\kappa^2}{\beta} \tag{1.5-13}$$

The corresponding formula for leaky TM modes is

$$2\alpha = \frac{2}{d}\left[\frac{n_2^2 \rho}{(n_2^2\kappa + n_1^2\rho)^2} + \frac{n_3^2(\Delta + \Delta^*)}{2|n_3^2\kappa + n_1^2\Delta|^2}\right]\frac{n_1^2\kappa^2}{\beta} \tag{1.5-14}$$

These equations can be expected to be reasonably accurate for relatively low losses. For very lossy waves the plane wave concept does not work very well.

It is important to realize that the leaky waves described so far are not modes of the structure. By definition a mode maintains its shape at all points along the cross section of the waveguide except for a phase change and except for an attenuation term of the form $e^{-\alpha z}$ in case of lossy modes. The lossy waves, whose properties we have studied by means of geometrical optics, are transient phenomena. If we connect a lossless waveguide to a guide whose dimensions are designed so that it is cutoff for the field distribution traveling in the first guide, we excite a field that behaves approximately as described by the ray optics theory of this section. However, this wave changes its shape throughout the cross section. This change of the field shape is caused by the radiation that is leaking out of the core region. Modes of dielectric waveguides must be considered throughout the infinite cross section. Outside the waveguide core we find more radiation at increasing distances from the core as we follow the leaky wave in the cutoff waveguide. We know that the true modes with propagation constants in the range (1.5-1) are radiation modes discussed in Sect. 1.4. Our leaky waves can be represented as series expansions in terms of radiation modes. Each radiation mode is lossless by itself. However, the total field must also be lossless because the power is simply redistributed from inside the waveguide core to the core region and the region outside the core. The loss

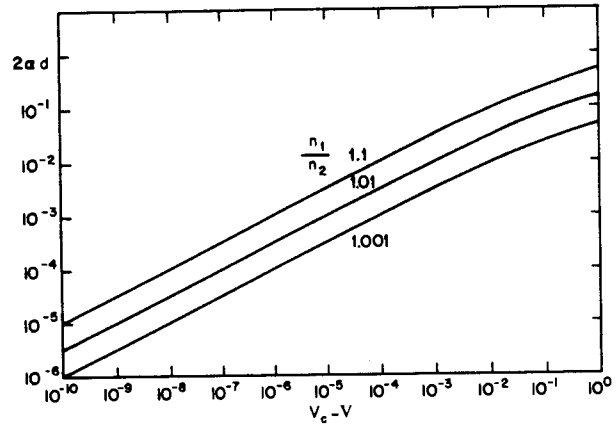

Fig. 1.5.2 *Power loss 2α of a leaky wave as a function of the frequency parameter V.*

formulas (1.5-13) and (1.5-14) account approximately for the power loss from the core region, regarding the amount of power found outside the core as lost.

Figure 1.5-2 shows a plot of the power loss coefficient (1.5-13), normalized with respect to the core width, as a function of the V parameter defined by Eq. (1.3-37). We have assumed $n_2 = n_3$ for simplicity. The curve applies to the second even TE mode, whose cutoff value is $V = V_c = 2\pi$. Since the loss increases enormously for V values just beyond cutoff, we have plotted $V_c - V$ on the horizontal axis. In order to have a feeling for the actual magnitude of the loss let us assume that the core is 5 μm wide. This corresponds approximately to a symmetric waveguide with $n_1/n_2 = 1.01$ and a vacuum wavelength of $\lambda = 1$ μm. In order to find the loss per centimeter expressed in decibels, we must divide the numbers given in Fig. 1.5.2 by the core width, $d = 5 \times 10^{-4}$ cm, and multiply the result by 4.34 to convert the absolute units to decibels. For $V_c - V = 10^{-8}$, we find for $n_1/n_2 = 1.01$ the loss value $2\alpha d = 3 \times 10^{-5}$ from Fig. 1.5.2. For our example of a 5-μm core, we thus have an actual loss of 0.26 dB/cm. This loss value increases by one order of magnitude for every two orders of magnitude increase of $V_c - V$. This means that for $V_c - V = 10^{-4}$, we already have 26 dB/cm loss. For $V_c - V = 0.1$, the loss has reached 651 dB/cm. The increase in loss levels off slightly for large values of $V_c - V$. This example dramatizes the very high losses of the leaky waves just beyond the cutoff point of the corresponding guided mode. The power that is carried in the core region escapes very fast once total internal reflection is lost.

After obtaining some insight into leaky waves by means of a geometrical optics approach, we continue their study from the point of view of the boundary

value approach. We have defined a guided mode as a solution of Maxwell's equations satisfying the boundary conditions at the core boundary and vanishing at $x = \pm\infty$. By omitting our insistence on vanishing field amplitudes at infinity, we obtained the radiation mode solutions of our boundary value problem. The radiation exhibited a continuous spectrum of allowed values of their propagation constants since we arrived at an inhomogeneous equations system for the determination of the amplitude coefficients of the field functions. This fundamental difference between guided and radiation modes was achieved by allowing incoming as well as outgoing waves to appear outside the waveguide core. The incoming waves add their own amplitude coefficients to the equation system, so that we end up with more undetermined coefficients than equations to determine them. This results in an inhomogeneous equation system if we arbitrarily assume that one of the amplitude coefficients is known. The equation system can then always be solved without the need for requiring the vanishing of the system determinant. There is, therefore, no determinantal or eigenvalue equation to restrict the possible values of the propagation constant.

However, it is possible to drop the requirement of vanishing field amplitudes at infinity without introducing additional fields corresponding to incoming waves, as in the case of radiation modes. We could, for example, search for solutions of the eigenvalue equation for negative instead of positive values of the parameters γ and δ which determine the amount of field decay outside the core region. If such solutions did exist, we would not expect them to correspond to physically realizable fields, since their field amplitudes would grow exponentially outside the core region. It turns out that solutions of the eigenvalue equations with complex values of γ and δ do indeed exist. For simplicity we study such solutions for the case of the symmetric slab waveguide. The even TE modes are determined by the eigenvalue equation

$$\tan u = v/u \tag{1.5-15}$$

and the odd modes by

$$\tan u = -u/v \tag{1.5-16}$$

These equations are identical with Eqs. (1.3-39) and (1.3-40). However, we are now using the abbreviations

$$V_s = \tfrac{1}{2}V = (n_1^2 - n_2^2)^{1/2}(kd/2) \tag{1.5-17}$$

$$u = \kappa d/2 \tag{1.5-18}$$

and

$$v = \gamma d/2 = (V_s^2 - u^2)^{1/2} \tag{1.5-19}$$

We discuss the leaky wave solutions of the even mode eigenvalue Eq. (1.5-15) in detail and state the corresponding results for the odd modes.

We begin by exploring the possibility of finding solutions of Eq. (1.5-15) for negative values of v. We set

$$v = -w, \qquad \text{with} \quad w > 0 \tag{1.5-20}$$

and search for solutions of the transcendental equation

$$\tan u = -w/u \tag{1.5-21}$$

Figure 1.5.3 gives an indication of the possible solutions for real positive values of w. The curve $-w/u$ was drawn under the assumption that V_s is slightly less than π, so that the second even mode is just beyond cutoff, according to Eqs. (1.3-43) and (1.5-17). The construction for the usual guided

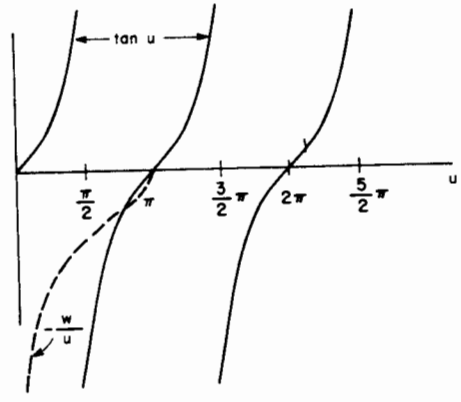

Fig. 1.5.3 *Graphical illustration of the unphysical solutions of the eigenvalue Eq. (1.5-21).*

modes differs from Fig. 1.5.3 by the fact that the dashed line representing the function v/u has the opposite sign and is folded upward to the first quadrant of the diagram. It is apparent from Fig. 1.5.3 that the dashed line crosses the tangent curves twice, provided that V_s is slightly smaller than π. There is, of course, also one crossing of the dashed curve with the second branch of the tangent curve for values $V_s > \pi$. Here we are in the regime where guided mode solutions of the second even mode also exist. The solutions of Eq. (1.5-21) for $V_s > \pi$ represent unphysical waves whose amplitudes grow with increasing distance from the core. The leaky wave solutions, that are apparent from the construction of Fig. 1.5.3, are also unphysical but merge continuously into the complex leaky wave solutions for which some physical rationale can be given. We are interested primarily in the solution that results from the crossover of the solid and dashed curve nearest the u axis. We see immediately that this unphysical solution starts at $u = \pi$ when $V_s = \pi$, that is, just at the cutoff

point of the second guided even TE mode. For decreasing values of V_s the crossover point moves slowly away from the u axis and yields decreasing values of the solution u. Then comes the point where the dashed curve becomes tangential to the tangent curve touching it in only one point. For even smaller values of V_s there are no more real solutions. We shall see shortly that Eq. (1.5-21) has complex solutions for values of V_s that are too small to yield a crossing of the solid and dashed curves of Fig. 1.5.3.

The value at which the real unphysical solution of Eq. (1.5-21) just ceases to exist is of interest. We obtain the solutions at this point by requiring that the tangents of the functions $\tan u$ and $-w/u$ are parallel and that Eq. (1.5-21) is still satisfied. The requirement of parallel tangents leads to the equation

$$1/\cos^2 u = V_s^2/wu^2 \tag{1.5-22}$$

From Eq. (1.5-21) we derive

$$\cos u = -u/V_s \tag{1.5-23}$$

The negative sign of the right-hand side of Eq. (1.5-23) follows from inspection of the sign of the cosine function for $\pi/2 < u < \pi$ and the fact that u and V_s must both be positive. Equations (1.5-22) and (1.5-23) must both be satisfied simultaneously. This is possible only if

$$w = 1 \tag{1.5-24}$$

The corresponding value of V_s is obtained by solving the transcendental Eq. (1.5-21) with $w = 1$:

$$\tan u = -1/u, \qquad \text{for even modes} \tag{1.5-25}$$

We have thus found the point where real solutions of Eq. (1.5-21) cease to exist. When we lower the value of V_s below the limit that is obtained from solutions of Eq. (1.5-25) and the relation

$$u = (V_s^2 - 1)^{1/2} \tag{1.5-25a}$$

the solutions of Eq. (1.5-21) or (1.5-15), if they exist at all, must be complex. We expect that the solutions move continuously from real to complex values, so that the imaginary parts of u and v must be small. We set

$$u = u_r + iu_i \tag{1.5-26}$$

and

$$v = v_r + iv_i \tag{1.5-27}$$

and require that

$$u_i \ll u_r, \qquad v_i \ll v_r \tag{1.5-28}$$

Substitution of Eqs. (1.5-26) and (1.5-27) into Eq. (1.5-15) leads to two simultaneous equations which assume the form

$$\tan u_r = v_r/u_r \tag{1.5-29}$$

and

$$u_i V_s^2/u_r^2 = v_i/u_r - u_i v_r/u_r^2 \tag{1.5-30}$$

if we use Eq. (1.5-28) and neglect terms of higher than second order in the small imaginary quantities. Second-order terms cancel from Eq. (1.5-29) if we use the fact that v_r can deviate from -1 only by a quantity of first order. Even though second-order terms are absent from these equations, they are indeed accurate to the second order of approximation. We need two more equations to be able to determine the four unknown quantities u_r, u_i, v_r, and v_i. These additional equations are obtained from Eq. (1.5-19). We find without any approximation

$$u_r^2 + v_r^2 - u_i^2 - v_i^2 = V_s^2 \tag{1.5-31}$$

and

$$u_r u_i + v_r v_i = 0 \tag{1.5-32}$$

When we solve Eq. (1.5-32) for v_i and substitute this value into Eq. (1.5-30) we obtain the solution

$$v_r = -1 \tag{1.5-33}$$

in agreement with Eq. (1.5-24). This means that the value v_r remains constant to first order of approximation and does not change from its value $v_r = -1$ which it assumes at the point where the solution becomes complex. Equation (1.5-29) results again in Eq. (1.5-25), which now applies to the real part u_r. Since u_r and v_r have been determined, we find the imaginary part v_i from Eqs. (1.5-31) and (1.5-32):

$$v_i = (u_r/V_s)(u_r^2 + 1 - V_s^2)^{1/2} \tag{1.5-34}$$

and also

$$u_i = (1/V_s)(u_r^2 + 1 - V_s^2)^{1/2} \tag{1.5-35}$$

We have thus found approximate leaky wave solutions of the eigenvalue Eq. (1.5-15) for the even TE modes of the symmetric slab waveguide. A complete solution must be obtained numerically with the aid of a computer. It may be helpful to point out that computer solutions tend to fail in the region where our approximate solutions are valid. The reason for this difficulty becomes apparent when we inspect the approximate forms (1.5-29) and (1.5-30) of the complex Eq. (1.5-15). The real part of the eigenvalue equation contains the zero-order terms u_r and v_r, while the imaginary part leads to Eq.

(1.5-30), which is small of first order. A computer solution of Eq. (1.5-15), which treats the imaginary and real parts of this complex equation on an equal footing, falls into the trap of being satisfied with small values of the imaginary part, regarding them as a sign that a solution is reached. In actuality, the imaginary part is always small near the point where the complex solution merges into the real solution, so that considerable inaccuracies result.

Figure 1.5.4 shows the locus of the solution points in the complex v plane. The parameter V_s is changed continuously. At the point $v_r = -1$, $v_i = 0$, we have V_s given as the solution of Eqs. (1.5-25) and (1.5-25a). As we move away from this point, V_s decreases in value and reaches zero when v_r reaches $-\infty$. The solution curves show the trajectories of the second even and odd modes. The first even mode does not have a cutoff value, so that it never becomes a leaky wave. The first odd mode, which does go through a cutoff at $V_s = \pi/2$, never leaves the negative real v axis and does not become a proper leaky wave. It simply does not have a physical interpretation beyond its cutoff point.

Even though our derivation was based on the eigenvalue Eq. (1.5-15) of the even TE modes, it can be shown that the approximate solutions (1.5-33)–

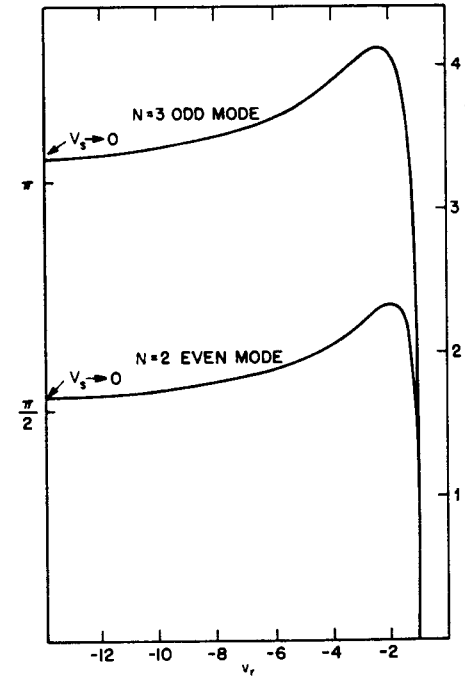

Fig. 1.5.4 *Trajectory of the leaky wave solutions in the complex v plane. The solution point moves as a function of V_s.*

(1.5-35) apply also to the odd TE modes, whose eigenvalue equation is Eq. (1.5-16). However, the real part of u is now obtained as the solution of the equation

$$\tan u_r = u_r, \qquad \text{for odd modes} \tag{1.5-36}$$

A similar analysis near the point $V_s = 0$ yields the approximate equations

$$\tan u_r = -u_i/u_r, \qquad \text{for even modes} \tag{1.5-37}$$

and

$$\tan u_r = u_r/u_i, \qquad \text{for odd modes} \tag{1.5-38}$$

and the equation

$$\exp(-u_i) = V_s/2(u_r^2 + u_i^2)^{1/2} \tag{1.5-39}$$

which holds for even and odd modes. Equation (1.5-39) shows that we must have

$$u_i \to \infty, \qquad \text{as} \quad V_s \to 0 \tag{1.5-40}$$

so that we find from Eqs. (1.5-37) and (1.5-38) the asymptotic end points of the trajectories shown in Fig. 1.5.4

$$u_r = v\pi/2 \tag{1.5-41}$$

with odd integer values of v for even modes and even integer values of v for odd modes. For $V_s = 0$ we find from Eq. (1.5-19)

$$v_i = u_r \tag{1.5-42}$$

and

$$v_r = -u_i \tag{1.5-43}$$

so that we also have

$$v_i = v\pi/2 \tag{1.5-43a}$$

This equation explains the asymptotic end points that are marked in Fig. 1.5.4.

Figure 1.5.5 shows schematically the general behavior of the trajectories of the leaky wave solutions as the parameter V_s runs through the values from 0 to ∞. Along the positive real v axis are the locations of the guided TE modes of the symmetric slab. For $V_s \to \infty$, the guided mode solutions move out to $v_r = \infty$. As the value of V_s decreases, each mode approaches its cutoff point at $v_r = 0$. One should think that any solution that exists past the cutoff point would immediately belong to a lossy wave with a positive imaginary part v_i. Positive values of v_i are needed since the wave should travel outward, away

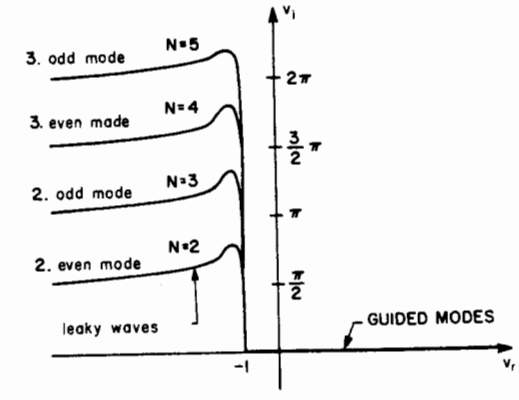

Fig. 1.5.5 *Schematic representation of the leaky wave solutions of the first four modes in the complex v plane.*

from the core. However, as we have discussed earlier and as is shown in Fig. 1.5.5, the locus of the solution of the eigenvalue equation first moves along the negative real v axis to the point $v_r = -1$. Starting at this point the solution moves out into the complex plane and can now be interpreted as a leaky mode corresponding to a wave that travels in the waveguide beyond cutoff. However, the field amplitude grows exponentially as $\exp[-(2v_r x/d)]$ for increasing values of x. We may interpret this behavior as the result of an accumulated radiation field that builds up from radiation escaping from the waveguide. Since the field must be visualized as arriving from $z = -\infty$, it appears plausible that the accumulated radiation would reach infinite values far from the core at $x = \pm\infty$, particularily since it requires an infinitely strong field at $z = -\infty$ if the field is to reach finite z values with a nonvanishing amplitude. It is thus possible to associate a certain physical reality with leaky wave solutions with $v_i > 0$. However, there are also complex solutions of the eigenvalue equation with $v_i < 0$ which correspond to waves that move inward toward the core and grow as they travel along the positive z axis. These waves and the solutions with $v_r < 0$ and $v_i = 0$ are of no particular practical value.

For small values of v_i, we can easily calculate the loss coefficient of the leaky wave. We consider a complex propagation constant

$$\beta = \beta_r - i\alpha \tag{1.5-44}$$

and obtain from Eqs. (1.2-14) and (1.5-19)

$$\beta_r = [n_2^2 k^2 + (2v_r/d)^2]^{1/2}, \qquad \text{for} \quad v_i \ll 1 \tag{1.5-45}$$

$$\alpha = -4v_r v_i/\beta_r d^2, \qquad \text{for} \quad v_i \ll 1 \tag{1.5-46}$$

We observe an interesting behavior of the real part of the propagation constant β of the leaky mode solution. As the wave approaches cutoff, v_r decreases in value and reaches $v_r = 0$ right at cutoff. At this point we have, not surprisingly, $\beta_r = n_2 k$. However, we should now expect that $\beta_r < n_2 k$ as the wave goes beyond cutoff. Instead, v_r becomes negative but stays real, so that we have again $\beta_r > n_2 k$ quite contrary to our expectations and to the behavior of the leaky wave transient fields discussed earlier in this section. In fact, as long as v_i remains small, we have for the real part of the propagation constant of the leaky mode solution

$$\beta_r = [n_2^2 k^2 + (4/d^2)]^{1/2} \tag{1.5-47}$$

independent of the mode number. If we do not neglect v_i we obtain

$$\beta_r = [n_2^2 k^2 + (v_r^2 - v_i^2)\, 4/d^2]^{1/2} \tag{1.5-48}$$

Using Eqs. (1.5-17) and (1.5-31) we can write this expression also in the form

$$\beta_r = [n_1^2 k^2 + (u_i^2 - u_r^2)\, 4/d^2]^{1/2} \tag{1.5-49}$$

For a mode of high order we have $u_r \gg 1$ and obtain from Eq. (1.5-29) with $v_r = -1$ and from Eq. (1.5-36) the approximation

$$u_r = N\pi/2 \tag{1.5-50}$$

Neglecting u_i we have from Eq. (1.5-49)

$$\beta_r = [n_1^2 k^2 - (N\pi/d)^2]^{1/2} \tag{1.5-51}$$

with even integer N for even modes and odd N for odd modes. This expression agrees with the β value given by Eqs. (1.5-2) and (1.5-3) which we deduced from our intuitive treatment of leaky waves.

The loss coefficient of Eq. (1.5-46) can be written in the following form if we use $v_r = -1$ and set $u_r \approx V_s \gg 1$ in Eq. (1.5-34):

$$\alpha = (4/\beta d^2)[(N\pi/2)^2 - V_s^2]^{1/2} \tag{1.5-52}$$

This approximation also holds only for high-order modes, so that we could replace u_r with $N\pi/2$ according to Eq. (1.5-50). In order to be able to compare the loss coefficient (1.5-52) with our earlier formula (1.5-13) we must first introduce the specialization to a symmetric slab waveguide. Setting $\Delta = \rho$ and using the fact that for a high-order, relatively low-loss mode we must have $\kappa \gg \rho$, we obtain from Eq. (1.5-13)

$$\alpha = 2\rho/\beta d \tag{1.5-53}$$

Next we replace β in Eq. (1.5-5) with expression (1.5-3) and use Eq. (1.5-17) with the result

$$\alpha = (4/\beta d^2)[(N\pi/2)^2 - V_s^2]^{1/2} \tag{1.5-54}$$

We thus see once more that for high-order modes our intuitive theory agrees with the result obtained from the leaky mode solutions of the eigenvalue equation. This agreement justifies our interpretation of the leaky mode solutions as belonging to waves that propagate beyond cutoff. It does seem surprising that we get this agreement. The leaky mode solutions are actually quite unphysical since their field amplitudes grow with increasing distance from the waveguide core. As we discussed earlier, power that is radiated from the core is actually not lost since it is simply redistributed from the core into the outside region. One wonders why this redistribution of power should appear as loss in a theory that treats the field throughout all infinite space as belonging to one and the same mode. It would not have been surprising if we had had no loss at all. In fact, our leaky mode does not show any loss just beyond the cutoff point. Only some distance (in V_s) beyond cutoff does the propagation constant assume complex values. The loss must thus indicate power outflow towards infinity in transverse direction. That the loss values of the leaky mode solution indeed agree with the loss calculation, which directly consideres the power outflow from the core, is not a trivial result. The conclusion to be reached from this discussion is that one should regard the loss coefficient obtained from the complex solutions of the eigenvalue equation with some reservation, since it is not entirely clear what constitutes loss in this theory. The intuitive formula (1.5-13) was obtained from the idea of counting all energy as lost once it escapes from the core. The intuitive theory does not actually describe a mode but a transient field condition. However, this is the situation one encounters as a wave enters a waveguide section where it is cutoff. The formula (1.5-13) must thus be considered more reliable for such applications.

1.6 Hollow Dielectric Waveguides

The leaky waves and leaky mode solutions discussed in the previous section occurred for waves that went beyond cutoff in a waveguide capable of supporting guided modes. We now consider a dielectric slab waveguide that does not support any guided mode in the proper sense of the word. So far we have always assumed that the slab waveguide of Fig. 1.1.1 had a core whose refractive index n_1 was larger than that of the surrounding media. If we invert the situation and assume that

$$n_1 < n_3 \leqslant n_2 \tag{1.6-1}$$

we obtain a structure that we call a hollow dielectric waveguide [MS1]. The name suggests that the core region may be air or vacuum with $n_1 = 1$. We thus think of a hollow channel in a dielectric medium. It is, of course, not

necessary that $n_1 = 1$; the type of leaky waveguide to be discussed here requires only that relationship (1.6-1) hold. According to Snell's law, Eq. (1.2-2), there can be no total internal reflection in this case. Every ray leaves the core with an increased angle after penetrating the core boundary.

Hollow dielectric waveguides have been successfully used in laser resonator structures [Sh1]. In this application a gas is confined inside a glass or quartz tube and is induced to exhibit gain by means of an electrical discharge. Gas lasers show increasing gain for decreasing tube diameter. The gain depends on the mechanism that depopulates the lower energy level of the atomic or molecular system that is used in the laser. This depopulation of the lower laser level is accomplished by collisions of the molecules with the wall. In a narrow tube these collisions occur more frequently, thus increasing the gain. However, as the tube diameter is decreased, the walls of the tube begin to interact with the electromagnetic field in the cavity. In a typical laser, the cavity is formed by external mirrors. The laser tube only has the purpose of confining the gas mixture. As the tube becomes too narrow its walls take part in shaping the mode field of the laser. This effect could be harmful if we think in terms of an open laser resonator defined solely by the mirrors. However, the tube walls can be used to advantage to guide the electromagnetic wave and thus confine not only the gas mixture but also the fields. The fact that a hollow dielectric waveguide does not actually have guided modes does not preclude its usefulness as a waveguide in this laser application. We have seen in the previous section that leaky waves propagate in dielectric waveguides for some distance. They may suffer high losses, but the amount of loss depends on the particular geometry and operating conditions that are being used.

The field of a hollow dielectric waveguide is also described by Eqs. (1.3-16)–(1.3-26) for TE modes and by Eqs. (1.3-54)–(1.3-63) for TM modes. The eigenvalue equations have the same form: Eq. (1.3-26) for TE modes and Eq. (1.3-63) for TM modes. The only difference is that now γ and δ defined by Eqs. (1.2-14) and (1.2-15) are imaginary for real values of β. There are thus no real solutions of the eigenvalue equations.

We obtain a very useful description of the leaky waves in the hollow dielectric waveguide by the geometrical optics approach that was used to describe leaky waves in the usual dielectric waveguide. If we ignore the eigenvalue equation and determine the real and imaginary parts of the propagation constants simply from the laws of ray optics and plane wave refraction and reflection at a dielectric interface, we obtain again Eqs. (1.5-2) and (1.5-3) for the real part of the propagation constant and Eq. (1.5-13) for twice its imaginary part, which is the power loss coefficient.

However, there is a difference in the behavior of the loss of leaky waves of ordinary waveguides compared to the leaky waves of hollow dielectric waveguides. In ordinary dielectric waveguides with $n_1 > n_2 \geqslant n_3$, we can assume

that either ρ or Δ of Eqs. (1.5-5) and (1.5-6) is very nearly zero, so that we have $\kappa \gg \rho$. For symmetric waveguides this assumption leads to the useful approximate form (1.5-53) for the loss coefficient. In low-loss hollow dielectric waveguides, we must assume that the ray impinges on the core boundary with very small angles, so that κ, given by Eq. (1.2-13), is very nearly zero. On the other side of the core boundary the angles are larger, so that we have

$$\kappa \ll \rho \tag{1.6-2}$$

and

$$\kappa \ll \Delta \tag{1.6-3}$$

A useful formula for the power loss coefficient 2α is thus obtained from Eq. (1.5-13):

$$2\alpha = (2\kappa^2/\beta d)(\rho^{-1} + \Delta^{-1}) \tag{1.6-4}$$

The allowed values of κ are given by Eq. (1.5-2):

$$\kappa = N\pi/d \tag{1.6-5}$$

so that we obtain from Eq. (1.5-5)

$$\rho = [(W/d)^2 + (N\pi/d)^2]^{1/2} \tag{1.6-6}$$

and from Eq. (1.5-6)

$$\Delta = [(n_3^2 - n_2^2)k^2 + (W/d)^2 + (N\pi/d)^2]^{1/2} \tag{1.6-7}$$

with the abbreviation

$$W = (n_2^2 - n_1^2)^{1/2} kd \tag{1.6-8}$$

Finally, we have for the propagation constant

$$\beta = [n_1^2 k^2 - (N\pi/d)^2]^{1/2} \tag{1.6-9}$$

In order to achieve low leaky wave loss in the hollow dielectric waveguide we must choose the mode number N small and the core width d large. In addition, the refractive index difference $n_3 - n_1$ must be made as large as possible. Low-loss propagation in a hollow dielectric waveguide is thus possible only for low-order modes in a guide that is capable of supporting very many leaky modes.

Let us consider an example: We use the free space wavelength $\lambda = 1\ \mu$m, $d = 50\ \mu$m, $n_1 = 1$, $n_2 = n_3 = 1.5$, and consider the lowest-order mode with $N = 1$. With these values, we obtain from Eq. (1.6-4)

$$2\alpha = 4.77 \times 10^{-2}\ \text{cm}^{-1} = 0.207\ \text{dB/cm} \tag{1.6-10}$$

This loss is moderately low. It is far too high for a waveguide for long distance communications. But it is low enough to be useful for short laser cavities or

integrated optics applications. The loss figure can, of course, be reduced to arbitrarily low values if we use larger values of d. However, we pay a penalty for trying to reduce the mode loss by increasing the core width. The waveguide becomes more vulnerable to losses caused by waveguide curvature. This effect is discussed in [MS1].

The leaky modes of the hollow dielectric waveguide do not have cutoff conditions since they are not properly guided in the usual sense. We obtain the total number of guided modes traveling in the forward direction by counting all the modes with positive propagation constant β. We use the symbol M to indicate the total number of forward traveling modes (we are talking only of TE modes for simplicity) and obtain $N = M$ from Eq. (1.6-9) by requiring that $\beta = 0$:

$$M = n_1 kd/\pi \tag{1.6-11}$$

For the condition used in our loss example we find from Eq. (1.6-11) that $M = 150$.

We conclude our discussion of hollow dielectric waveguides by considering the solutions of the eigenvalue Eq. (1.3-39)

$$\tan(\kappa d/2) = \gamma/\kappa \tag{1.6-12}$$

for the even TE modes of the symmetric hollow slab waveguide with $n_2 = n_3$. The x dependence of the field is given by [see Eqs. (1.3-16) and (1.3-18) with $\gamma = \delta$]

$$e^{-\gamma x} \tag{1.6-13}$$

We must require that the imaginary part of γ is positive in order to assure that the waves move outward, away from the core. The difference between our present solution and the solution discussed in the previous section is that, according to Eq. (1.6-2), the imaginary part of γ must now be much larger than its real part because the imaginary part corresponds to ρ, while the real part is even smaller than κ. We again set

$$\kappa d/2 = u_r + iu_i \tag{1.6-14}$$

and

$$\gamma d/2 = v_r + iv_i \tag{1.6-15}$$

and require that

$$u_r \gg u_i \tag{1.6-16}$$

and

$$v_i \gg v_r \tag{1.6-17}$$

Substitution of Eqs. (1.6-14) and (1.6-15) into Eq. (1.6-12) and use of the inequalities (1.6-16) and (1.6-17) yields the two simultaneous equations

$$[\tan u_r(1-\tanh^2 u_i)]/(1+\tan^2 u_r \tanh^2 u_i) = v_i u_i/u_r^2 \tag{1.6-18}$$

and

$$[\tanh u_i(1+\tan^2 u_r)]/(1+\tan^2 u_r \tanh^2 u_i) = v_i/u_r \tag{1.6-19}$$

The right-hand side of Eq. (1.6-19) is very large. If we assume that $\tan u_r \gg 1$ and $\tan u_r \tanh u_i \gg 1$, we obtain from Eq. (1.6-19)

$$1/u_i = v_i/u_r \tag{1.6-20}$$

provided that u_i itself is small. With the same assumptions we can write Eq. (1.6-18) approximately in the form

$$1/u_i^2 \tan u_r = v_i u_i/u_r^2 \tag{1.6-21}$$

Next we set (N = odd integer)

$$u_r = N\pi/2 - \eta \tag{1.6-22}$$

and obtain from Eq. (1.6-21)

$$\eta/u_i^2 = v_i u_i/(N\pi/2)^2 \tag{1.6-23}$$

and from Eq. (1.6-20)

$$u_i v_i = N\pi/2 \tag{1.6-24}$$

The relation

$$\kappa^2 + \gamma^2 = (n_1^2 - n_2^2)k^2 \tag{1.6-25}$$

that follows from Eqs. (1.2-13) and (1.2-14) leads, with our present assumptions, to the two relations

$$v_i = [W_s^2 + (N\pi/2)^2]^{1/2} \tag{1.6-26}$$

and

$$v_r = -(u_i/v_i)\, N\pi/2 \tag{1.6-27}$$

where we have used the abbreviation

$$W_s = [n_2^2 - n_1^2]^{1/2} kd/2 \tag{1.6-28}$$

We now obtain immediately from Eq. (1.6-24)

$$u_i = \frac{N\pi/2}{[W_s^2 + (N\pi/2)^2]^{1/2}} \tag{1.6-29}$$

and from Eqs. (1.6-23), (1.6-24), and (1.6-29)

$$\eta = \frac{N\pi/2}{W_s^2 + (N\pi/2)^2} \tag{1.6-30}$$

Equation (1.6-27) finally yields

$$v_r = \frac{-(N\pi/2)^2}{W_s^2 + (N\pi/2)^2} \tag{1.6-31}$$

Our approximation is apparently valid for $W_s \gg 1$.

It is noteworthy that this approximate solution implies that the leaky waves of the hollow dielectric waveguide also grow in transverse direction away from the core since v_r is negative, so that Eq. (1.6-13) is an exponentially growing function.

We obtain the propagation constant from the equation

$$\beta_r - i\alpha = (n_1^2 k^2 - \kappa^2)^{1/2} \tag{1.6-32}$$

so that we have for its real part

$$\beta_r = [n_1^2 k^2 - (N\pi/d)^2]^{1/2} \tag{1.6-33}$$

and for its imaginary part, which is the amplitude attenuation constant,

$$\alpha = \frac{u_r u_i}{\beta_r} \frac{4}{d^2} = \frac{4(N\pi/2)^2}{\beta_r [W_s^2 + (N\pi/2)^2]^{1/2} d^2} \tag{1.6-34}$$

Equation (1.6-33) is identical with Eq. (1.6-9) and the equation for the attenuation constant (1.6-34) is identical with (1.6-4) if we remember that for the symmetrical slab we must set $\rho = \Delta$ and $W = 2W_s$.

The agreement of the solution of the eigenvalue equation with the result of the geometrical optics treatment is much better for the hollow dielectric waveguide than it was for the leaky waves of the ordinary dielectric waveguide. In our present case we did not have to assume that we were dealing with a high-order mode. The two methods of approximate analysis agree for any mode, provided that the losses are low. Our results, Eqs. (1.6-26)–(1.6-34), apply also to the odd modes. It is interesting that we must now associate odd values of N with even modes and even N values with odd modes, contrary to the assignment that is listed below Eq. (1.5-51). The difference is caused by the fact that at the core boundary the fields of low-loss modes in the hollow dielectric waveguide are very small, while they are strong for the modes of the ordinary waveguide just below cutoff.

The trajectory of the solution points of the complex v values for $0 < W_s < \infty$ is shown schematically in Fig. 1.6.1. This figure was drawn with the help of numerical computer solutions of the eigenvalue equations for even

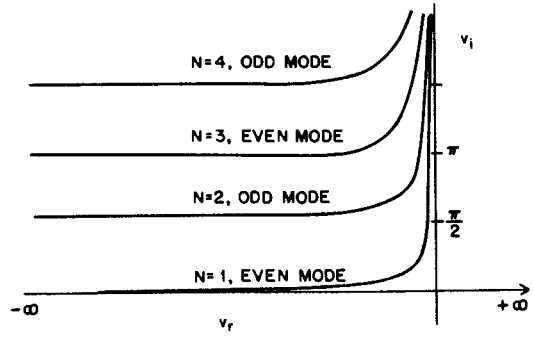

Fig. 1.6.1 *Leaky wave trajectories in the complex v plane for hollow dielectric waveguides.*

and odd modes of the symmetric hollow dielectric waveguide. For very large values of W_s, the trajectory of solutions points moves up along the positive imaginary v axis reaching $v_r = 0$ and $v_i = \infty$ at $W_s = \infty$. For decreasing values of W_s, the values of v_r become increasingly more negative, while v_i decreases. In the limit of $W_s = 0$, we have $v_r = -\infty$ and

$$v_i = v\pi/2, \qquad \text{for} \quad W_s = 0 \tag{1.6-35}$$

with even integer values of v for even modes and odd v values for odd modes. The assignment of even and odd integers to even and odd modes is again reversed compared to the corresponding situation for ordinary dielectric waveguides.

1.7 Rectangular Dielectric Waveguides

The dielectric slab waveguide is a useful model for more complicated waveguide structures. Its simplicity allows us to study the properties of wave propagation in dielectric waveguides without the encumberance of tedious mathematical expressions. However, in most practical applications more complicated waveguides are used. The waveguides used in integrated optics are usually rectangular strips of dielectric material that are embedded in other

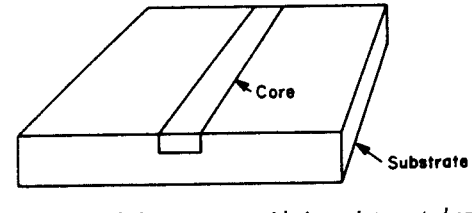

Fig. 1.7.1 *Rectangular dielectric waveguide in an integrated optics application.*

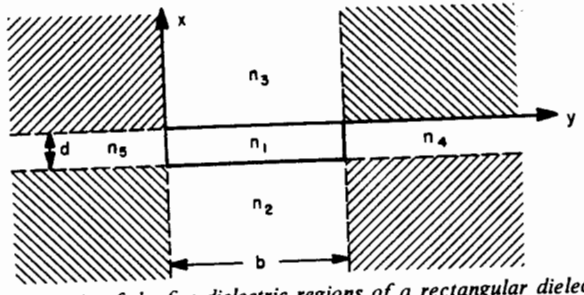

Fig. 1.7.2 *Schematic of the five dielectric regions of a rectangular dielectric waveguide. The field in the shaded regions is ignored in the approximation.*

dielectrics. Figure 1.7.1 shows the geometry of a rectangular dielectric waveguide. The rectangular strip is embedded in the material of the substrate of the integrated circuit. We analyze a structure that is more general: Instead of assuming that the waveguide core is embedded in the material of a substrate we allow the materials on all four sides of the rectangular core to be different. This geometry is shown in Fig. 1.7.2. We have chosen the labeling of the refractive indices to coincide as nearly as possible with Fig. 1.1.1.

An exact analytical treatment of this problem is not possible. Approximate solutions by numerical methods have been obtained that can be made as accurate as desired [GL1]. However, here we are following an approximate analytical approach that was developed by Marcatili [Mi1]. This method works only for modes far from cutoff. If the mode is not very close to cutoff, its field is confined almost exclusively to the region of the core, and only very little field energy is carried in the surrounding media. In particular, there is very little field energy in the shaded regions of Fig. 1.7.2, so that we can simplify the analysis by ignoring these regions completely. We can verify the statement that most of the field is contained in the core by examining the solutions of the slab waveguide. The ray angle θ_1 is small for well-guided modes, so that we see from Eq. (1.2-6) that the propagation constant approaches

$$\beta \approx n_1 k \tag{1.7-1}$$

Because k is a large quantity, we thus have from Eqs. (1.2-13)–(1.2-15)

$$\kappa \ll \gamma \tag{1.7-2}$$

and

$$\kappa \ll \delta \tag{1.7-3}$$

According to Eq. (1.3-25) we have $|B/A| \gg 1$, so that the sine term in the field expression (1.3-17) predominates over the cosine term. At the upper core boundary, $x = 0$, we find that the field is quite weak compared to its maximum value inside the core. Inequalities (1.7-2) and (1.7-3) also cause $\sin \kappa d$ of

Eq. (1.3-28) to be small. We thus see from Eq. (1.3-17) that the field is also weak at the lower core boundary, $x = -d$. The E_y field amplitude just outside the core is equal to the value on the inside and drops off with increasing distance from the core. It is thus clear that the field in the core is much stronger than the field outside as long as the mode is not too close to cutoff. The same considerations hold also for the TM mode [Eq. (1.3-55)].

It is possible to express the transverse field components in terms of the longitudinal components. With the time and z dependence, Eq. (1.3-8), we obtain from Maxwell's Eqs. (1.3-2) and (1.3-3) (compare p. 13 of [Me1])

$$E_x = -(i/K_j^2)[(\beta\, \partial E_z/\partial x) + (\omega\mu_0\, \partial H_z/\partial y)] \tag{1.7-4}$$

$$E_y = -(i/K_j^2)[(\beta\, \partial E_z/\partial y) - (\omega\mu_0\, \partial H_z/\partial x)] \tag{1.7-5}$$

$$H_x = -(i/K_j^2)[(\beta\, \partial H_z/\partial x) - (\omega n_j^2 \varepsilon_0\, \partial E_z/\partial y)] \tag{1.7-6}$$

$$H_y = -(i/K_j^2)[(\beta\, \partial H_z/\partial y) + (\omega n_j^2 \varepsilon_0\, \partial E_z/\partial x)] \tag{1.7-7}$$

The parameter K_j is defined by

$$K_j = (n_j^2 k^2 - \beta^2)^{1/2} \tag{1.7-8}$$

The refractive index n_j assumes the values n_1, n_2, n_3, n_4, and n_5 for fields in the five regions of the waveguide. The longitudinal components E_z and H_z must satisfy the reduced wave equation

$$(\partial^2\psi/\partial x^2) + (\partial^2\psi/\partial y^2) + K_j^2\psi = 0 \tag{1.7-9}$$

There are two types of modes that the waveguide can support. One type, which we call E^x_{pq} modes, is polarized predominantly in the x direction. The other mode, E^y_{pq}, is polarized predominantly in the y direction. We can adjust the amplitude coefficients of E_z and H_z so that one of the transverse field components vanishes. The following set of field components satisfies the reduced wave equation and relations (1.7-4)–(1.7-7) and describes an E^x_{pq} mode in region 1. The factor (1.3-8) is omitted throughout:

$$E_z = A \cos\kappa_x(x+\xi) \cos\kappa_y(y+\eta) \tag{1.7-10}$$

$$H_z = -A(\varepsilon_0/\mu_0)^{1/2} n_1^2 (\kappa_y/\kappa_x)(k/\beta) \sin\kappa_x(x+\xi) \sin\kappa_y(y+\eta) \tag{1.7-11}$$

$$E_x = (iA/\kappa_x\beta)(n_1^2 k^2 - \kappa_x^2) \sin\kappa_x(x+\xi) \cos\kappa_y(y+\eta) \tag{1.7-12}$$

$$E_y = -iA(\kappa_y/\beta) \cos\kappa_x(x+\xi) \sin\kappa_y(y+\eta) \tag{1.7-13}$$

$$H_x = 0 \tag{1.7-14}$$

$$H_y = iA(\varepsilon_0/\mu_0)^{1/2} n_1^2 (k/\kappa_x) \sin\kappa_x(x+\xi) \cos\kappa_y(y+\eta) \tag{1.7-15}$$

with $k = \omega^2\varepsilon_0\mu_0$ and

$$K_1^2 = n_1^2 k^2 - \beta^2 = \kappa_x^2 + \kappa_y^2 \tag{1.7-16}$$

The longitudinal components E_z and H_z were chosen to satisfy the reduced wave Eq. (1.7-9). This requirement leads to condition (1.7-16). Except for this restriction κ_x and κ_y are arbitrary at this point. The sine and cosine functions were selected so that the contributions of E_z and H_z result in the same functional dependence of the transverse components. The phase parameters ξ and η, which are added to the arguments of the sine and cosine functions, are necessary to make the solution general. We could have used a superposition of a sine and cosine function with arbitrary amplitude coefficients instead of using a circular function with an arbitrary amplitude and phase. Both methods are, of course, equivalent. The various undetermined coefficients become fixed by the requirement that the fields satisfy the boundary conditions on the four sides of the core.

The choice of the amplitude coefficient for H_z may appear arbitrary. However, it was dictated by the desire to make the H_y component vanish. Relation (1.7-1) results in the inequalities

$$\kappa_x \ll \beta \tag{1.7-17}$$

and

$$\kappa_y \ll \beta \tag{1.7-18}$$

which follow from Eq. (1.7-16) and apply to modes far from cutoff. Owing to inequality (1.7-17) we see that $E_y \ll E_z$. Since $E_z \ll E_x$ we can neglect E_y as a quantity of second order in κ/β. To the approximation used in this analysis we have thus only four nonvanishing field components.

On the outside we must require all field components in regions 2–5 to vanish at infinite distance from the core. We are thus limited to use decaying exponential functions for those coordinates that lead away from the core. The field components in these regions are chosen similarly to the field in the core, with the additional requirement that the electric field component tangential to the particular face of the core passes continuously through the core boundary. We thus have in region 2

$$E_z = A \cos\kappa_x(\xi - d) \cos\kappa_y(y+\eta) \exp[\gamma_2(x+d)] \tag{1.7-19}$$

$$H_z = -A(\varepsilon_0/\mu_0)^{1/2} n_2^2 (\kappa_y/\gamma_2)(k/\beta) \cos\kappa_x(\xi - d) \sin\kappa_y(y+\eta) \exp[\gamma_2(x+d)] \tag{1.7-20}$$

$$E_x = +iA[(\gamma_2^2 + n_2^2 k^2)/\gamma_2\beta] \cos\kappa_x(\xi - d) \cos\kappa_y(y+\eta) \exp[\gamma_2(x+d)] \tag{1.7-21}$$

$$E_y \approx 0 \tag{1.7-22}$$

$$H_x = 0 \tag{1.7-23}$$

$$H_y = iA(\varepsilon_0/\mu_0)^{1/2} n_2^2 (k/\gamma_2) \cos\kappa_x(\xi - d) \cos\kappa_y(y+\eta) \exp[\gamma_2(x+d)] \tag{1.7-24}$$

with

$$K_2^2 = n_2^2 k^2 - \beta^2 = \kappa_y^2 - \gamma_2^2 \tag{1.7-25}$$

The amplitude coefficients of these field components were chosen to ensure continuity of the E_z component at $x = -d$ and to make $H_x = 0$. The E_y component vanishes to the same approximation as the E_y component inside the core. The y-dependent functions were chosen to coincide with the corresponding functions inside the core. Equation (1.7-25) is again a consequence of Eq. (1.7-9).

The field in region 3 is similarly

$$E_z = A \cos\kappa_x\xi \cos\kappa_y(y+\eta) \exp(-\gamma_3 x) \tag{1.7-26}$$

$$H_z = A(\varepsilon_0/\mu_0)^{1/2} n_3^2 (\kappa_y/\gamma_3)(k/\beta) \cos\kappa_x\xi \sin\kappa_y(y+\eta) \exp(-\gamma_3 x) \tag{1.7-27}$$

$$E_x = -iA[(\gamma_3^2 + n_3^2 k^2)/\gamma_3\beta] \cos\kappa_x\xi \cos\kappa_y(y+\eta) \exp(-\gamma_3 x) \tag{1.7-28}$$

$$E_y \approx 0 \tag{1.7-29}$$

$$H_x = 0 \tag{1.7-30}$$

$$H_y = -iA(\varepsilon_0/\mu_0)^{1/2} n_3^2 (k/\gamma_3) \cos\kappa_x\xi \cos\kappa_y(y+\eta) \exp(-\gamma_3 x) \tag{1.7-31}$$

with

$$K_3^2 = n_3^2 k^2 - \beta^2 = \kappa_y^2 - \gamma_3^2 \tag{1.7-32}$$

In regions 4 and 5 we adjust the amplitudes so that the strong component E_x is continuous at the core boundary and obtain in region 4

$$E_z = A(n_1^2/n_4^2) \cos\kappa_y(b+\eta) \cos\kappa_x(x+\xi) \exp[-\gamma_4(y-b)] \tag{1.7-33}$$

$$H_z = -A(\varepsilon_0/\mu_0)^{1/2} n_1^2 (\gamma_4/\kappa_x)(k/\beta) \cos\kappa_y(b+\eta) \sin\kappa_x(x+\xi) \exp[-\gamma_4(y-b)] \tag{1.7-34}$$

$$E_x = iA(n_1^2/n_4^2)[(n_4^2 k^2 - \kappa_x^2)/\kappa_x\beta] \cos\kappa_y(b+\eta) \times \sin\kappa_x(x+\xi) \exp[-\gamma_4(y-b)] \tag{1.7-35}$$

$$H_y = iA(\varepsilon_0/\mu_0)^{1/2} n_1^2 (k/\kappa_x) \cos\kappa_y(b+\eta) \sin\kappa_x(x+\xi) \exp[-\gamma_4(y-b)] \tag{1.7-36}$$

with

$$K_4^2 = n_4^2 k^2 - \beta^2 = \kappa_x^2 - \gamma_4^2 \tag{1.7-37}$$

In region 5 we have finally

$$E_z = A(n_1^2/n_5^2)\cos\kappa_y\eta\cos\kappa_x(x+\xi)\exp(\gamma_5 y) \tag{1.7-38}$$

$$H_z = A(\varepsilon_0/\mu_0)^{1/2}n_1^2(\gamma_5/\kappa_x)(k/\beta)\cos\kappa_y\eta\sin\kappa_x(x+\xi)\exp(\gamma_5 y) \tag{1.7-39}$$

$$E_x = iA(n_1^2/n_5^2)[(n_5^2k^2-\kappa_x^2)/\kappa_x\beta]\cos\kappa_y\eta\sin\kappa_x(x+\xi)\exp(\gamma_5 y) \tag{1.7-40}$$

$$H_y = iA(\varepsilon_0/\mu_0)^{1/2}n_1^2(k/\kappa_x)\cos\kappa_y\eta\sin\kappa_x(x+\xi)\exp(\gamma_5 y) \tag{1.7-41}$$

with

$$K_5^2 = n_5^2k^2 - \beta^2 = \kappa_x^2 - \gamma_5^2 \tag{1.7-42}$$

It is noteworthy that, in the spirit of our approximation, the E_x component is continuous only if we neglect the κ_x^2 term in the numerator of the equations compared to the term n^2k^2. The H_x component vanishes exactly, while E_y is only approximately zero in regions 4 and 5.

We complete the analysis of the modes of the rectangular dielectric waveguide by matching the remaining boundary conditions. One tangential electric field component has already been matched by proper choice of the field amplitudes. In regions 2 and 3 we require that the H_z component pass continuously through the core boundary at $x = 0$ and $x = -d$. This requirement also causes H_y to be continuous as required. The remaining tangential component E_y, which would have to be continuous in an exact treatment, is neglected since it is small compared to all other field components. We thus obtain the following two equations:

$$(n_1^2/\kappa_x)\sin\kappa_x(\xi-d) - (n_2^2/\gamma_2)\cos\kappa_x(\xi-d) = 0 \tag{1.7-43}$$

and

$$(n_1^2/\kappa_x)\sin\kappa_x\xi + (n_3^2/\gamma_3)\cos\kappa_x\xi = 0 \tag{1.7-44}$$

If the sine and cosine functions in Eq. (1.7-43) are expanded by means of the addition theorems, these two equations represent a system of homogeneous simultaneous equations for the two unknowns, $\sin\kappa_x\xi$ and $\cos\kappa_x\xi$. A solution is possible only if the determinant of the equation system vanishes. We thus obtain the eigenvalue equation

$$\tan\kappa_x d = n_1^2\kappa_x(n_3^2\gamma_2 + n_2^2\gamma_3)/(n_3^2n_2^2\kappa_x^2 - n_1^4\gamma_2\gamma_3) \tag{1.7-45}$$

We recognize this equation as the eigenvalue equation of TM modes of the infinite slab, Eq. (1.3-63). It is clear that relative to the two core boundaries at $x = 0$ and $x = -d$ the field polarization corresponds to a TM mode, because the transverse electric field is polarized normal to the core boundary.

The phase parameter ξ is determined by Eq. (1.7-44):

$$\tan\kappa_x\xi = -(n_3^2/n_1^2)(\kappa_x/\gamma_3) \tag{1.7-46}$$

which corresponds to Eq. (1.3-61). We see that the eigenvalue Eq. (1.7-45) can be used to determine κ_x, since we can express γ_2 in terms of κ_x with the help of Eqs. (1.7-16) and (1.7-25):

$$\gamma_2 = [(n_1^2 - n_2^2)k^2 - \kappa_x^2]^{1/2} \tag{1.7-47}$$

Similarly it is possible to express γ_3 in terms of κ_x by means of Eqs. (1.7-16) and (1.7-32):

$$\gamma_3 = [(n_1^2 - n_3^2)k^2 - \kappa_x^2]^{1/2} \tag{1.7-48}$$

This leaves κ_x as the only unknown quantity in Eq. (1.7-45).

At the core boundaries in regions 4 and 5 we require that H_z assume equal values on either side of the dielectric interface. The E_x component is already continuous by our choice of wave amplitudes, provided we neglect κ_x^2 compared to n^2k^2. We obtain the following two equations from the requirement of continuity of H_z at $y = b$ and $y = 0$:

$$\kappa_y\sin\kappa_y(b+\eta) - \gamma_4\cos\kappa_y(b+\eta) = 0 \tag{1.7-49}$$

and

$$\kappa_y\sin\kappa_y\eta + \gamma_5\cos\kappa_y\eta = 0 \tag{1.7-50}$$

The E_z component is not continuous at these two interfaces. However, far from cutoff condition (1.7-2) indicates that $(\mu_0/\varepsilon_0)^{1/2}H_z$ of Eq. (1.7-34) is considerably larger than E_z of Eq. (1.7-33). Matching of the H_z component is thus more important than letting E_z satisfy the boundary conditions. In addition, it is obvious that the E_z component satisfies the boundary condition in the limit $n_1 = n_4 = n_5$. The approximation thus gets better for small index differences.

From Eq. (1.7-50) we find that the phase parameter η is given by the relation

$$\tan\kappa_y\eta = -\gamma_5/\kappa_y \tag{1.7-51}$$

In Eq. (1.7-49) we expand the sine and cosine functions by means of the addition theorems and obtain an equation system for the unknown quantities $\sin\kappa_y\eta$ and $\cos\kappa_y\eta$. The condition of vanishing system determinant of this equation system yields the eigenvalue equation

$$\tan\kappa_y b = \kappa_y(\gamma_4+\gamma_5)/(\kappa_y^2 - \gamma_4\gamma_5) \tag{1.7-52}$$

which we recognize as Eq. (1.3-26) for TE modes of the slab. It is also apparent that Eq. (1.7-51) corresponds to Eq. (1.3-25). This result is reasonable. E_x is the dominant electric field component of the E_{pq}^x modes, so that the field

appears as a TE mode when viewed from region 4 or 5. Equation (1.7-52) determines κ_y, since we obtain from Eqs. (1.7-16), (1.7-37), and (1.7-42)

$$\gamma_4 = [(n_1^2 - n_4^2)k^2 - \kappa_y^2]^{1/2} \tag{1.7-53}$$

and

$$\gamma_5 = [(n_1^2 - n_5^2)k^2 - \kappa_y^2]^{1/2} \tag{1.7-54}$$

Once κ_x and κ_y are determined, we obtain the propagation constant from Eq. (1.7-16):

$$\beta = [n_1^2 k^2 - (\kappa_x^2 + \kappa_y^2)]^{1/2} \tag{1.7-55}$$

This completes our determination of the E_{pq}^x mode. The integers p and q are the mode numbers which we obtain from the solutions of the two eigenvalue Eqs. (1.7-45) and (1.7-52). These numbers indicate the number of maxima of the field distribution in x and y direction. We have obtained a field description inside the waveguide core and in regions 2–5 as shown in Fig. 1.7.2. However, in the shaded regions of this figure the field remains unknown. Ignoring these areas makes it impossible to obtain an exact solution by this method. We have seen that we could not satisfy all the boundary conditions, even in the four unshaded regions, since our method is inherently not correct. However, we shall see later that our approximation compares favorably with more exact numerical solutions of the problem.

The E_{pq}^y modes are obtained in close analogy to the derivation of the E_{pq}^x modes. It is now the E_y component that is dominant, while the E_x component nearly vanishes, and $H_y = 0$. We restrict the description of the field components to stating only the E_z and H_z components in all four media. The corresponding transverse components are obtained by differentiation from Eqs. (1.7-4)–(1.7-7). We obtain in the core of the waveguide, that is, in medium 1,

$$E_z = B \cos \kappa_x (x + \bar{\xi}) \cos \kappa_y (y + \bar{\eta}) \tag{1.7-56}$$

$$H_z = B(\varepsilon_0/\mu_0)^{1/2} n_1^2 (\kappa_x/\kappa_y)(k/\beta) \sin \kappa_x (x + \bar{\xi}) \sin \kappa_y (y + \bar{\eta}) \tag{1.7-57}$$

In medium 2 we have

$$E_z = B(n_1^2/n_2^2) \cos \kappa_x (\bar{\xi} - d) \cos \kappa_y (y + \bar{\eta}) \exp[\gamma_2 (x + d)] \tag{1.7-58}$$

$$H_z = -B(\varepsilon_0/\mu_0)^{1/2} n_1^2 (\gamma_2/\kappa_y)(k/\beta) \cos \kappa_x (\bar{\xi} - d) \sin \kappa_y (y + \bar{\eta}) \exp[\gamma_2 (x + d)] \tag{1.7-59}$$

The following field components belong to medium 3:

$$E_z = B(n_1^2/n_3^2) \cos \kappa_x \bar{\xi} \cos \kappa_y (y + \bar{\eta}) \exp(-\gamma_3 x) \tag{1.7-60}$$

$$H_z = B(\varepsilon_0/\mu_0)^{1/2} n_1^2 (\gamma_3/\kappa_y)(k/\beta) \cos \kappa_x \bar{\xi} \sin \kappa_y (y + \bar{\eta}) \exp(-\gamma_3 x) \tag{1.7-61}$$

The field in medium 4 is described by

$$E_z = B \cos \kappa_y (b + \bar{\eta}) \cos \kappa_x (x + \bar{\xi}) \exp[-\gamma_4 (y - b)] \tag{1.7-62}$$

$$H_z = -B(\varepsilon_0/\mu_0)^{1/2} n_4^2 (\kappa_x/\gamma_4)(k/\beta) \cos \kappa_y (b + \bar{\eta}) \sin \kappa_x (x + \bar{\xi}) \exp[-\gamma_4 (y - b)] \tag{1.7-63}$$

Finally we have in medium 5:

$$E_z = B \cos \kappa_y \bar{\eta} \cos \kappa_x (x + \bar{\xi}) \exp(\gamma_5 y) \tag{1.7-64}$$

$$H_z = B(\varepsilon_0/\mu_0)^{1/2} n_5^2 (\kappa_x/\gamma_5)(k/\beta) \cos \kappa_y \bar{\eta} \sin \kappa_x (x + \bar{\xi}) \exp(\gamma_5 y) \tag{1.7-65}$$

Equations (1.7-16), (1.7-25), (1.7-32), (1.7-37), and (1.7-42) also apply to this case.

In regions 2 and 3 we require continuity of the H_z component at the core boundary and obtain in the usual way the eigenvalue equation for TE modes of a slab:

$$\tan \kappa_x d = \kappa_x (\gamma_2 + \gamma_3)/(\kappa_x^2 - \gamma_2 \gamma_3) \tag{1.7-66}$$

and the following equation for $\bar{\xi}$,

$$\tan \kappa_x \bar{\xi} = \gamma_3/\kappa_x \tag{1.7-67}$$

Continuity of E_y to first order in κ_y/k at these two interfaces is assured by our choice of the amplitude coefficients of E_z and H_z. The boundary condition for E_z is not satisfied. However, the magnitude of E_z is small compared to the other electric field components and, in addition, better compliance with this boundary condition is obtained for decreasing refractive index differences.

In regions 4 and 5 we match the H_z components on either side of the core boundary and obtain the eigenvalue equation for TM modes of the asymmetric slab:

$$\tan \kappa_y b = n_1^2 \kappa_y (n_5^2 \gamma_4 + n_4^2 \gamma_5)/(n_4^2 n_5^2 \kappa_y^2 - n_1^4 \gamma_4 \gamma_5) \tag{1.7-68}$$

and the equation for $\bar{\eta}$,

$$\tan \kappa_y \bar{\eta} = (n_5^2/n_1^2)(\kappa_y/\gamma_5) \tag{1.7-69}$$

The H_x component becomes continuous after we have matched the H_z component. Continuity of the E_z component is assured by our choice of the amplitude coefficients and the E_x component is neglected, since it is small to second order in κ_y/k compared to the E_y component. The two eigenvalue equations are again used to determine κ_x and κ_y, since γ_j can be expressed in terms of κ_x and κ_y as before. The propagation constant is finally obtained from Eq. (1.7-55).

Our approximate analysis does not hold near cutoff since the fields detach themselves from the core and reach strongly into the shaded regions of Fig.

1.7.2. This breakdown of the theory near cutoff is also apparent from the propagation constant. We write Eq. (1.7-55) in the form

$$\beta d = [n_2^2 k^2 d^2 + V^2 - (\kappa_x^2 + \kappa_y^2) d^2]^{1/2} \tag{1.7-70}$$

If n_2 is larger than any of the other refractive indices—except n_1—total internal reflection will first break down at the core boundary $x = -d$. According to Eq. (1.3-74) we have right at cutoff $V = V_c$ and

$$\kappa_c d = V_c . \tag{1.7-71}$$

so that the propagation constant assumes the value

$$\beta = (n_2^2 k^2 - \kappa_y^2)^{1/2} < n_2 k \tag{1.7-72}$$

Even if we assume $n_2 = n_3 = n_4 = n_5$, Eq. (1.7-72) implies that the propagation constant, at the point where the wave is no longer guided, is smaller than the plane wave propagation constant in the medium outside the waveguide core. This is an unphysical result which shows that our theory just does not work near the cutoff point, because when most of the field energy is traveling outside the core we must have $\beta = n_2 k$.

Marcatili [Mi1] has evaluated the results of the approximate field analysis and compared them to computer solutions of the problem by Goell. Of the many figures given in Marcatili's paper we reproduce two that apply to the case of a square guide and of a rectangular guide with $b = 2d$. In both cases we have $n_2 = n_3 = n_4 = n_5$. The solid curves in Fig. 1.7.3 were obtained from exact solutions of the eigenvalue Eqs. (1.7-45), (1.7-52), (1.7-66) and (1.7-68). The dashed curves show the result of the far from cutoff approximation (1.3-88) and (1.3-91) of these eigenvalue equations. However, Marcatili used the approximation

$$\delta_\infty = \gamma_\infty = (n_1^2 - n_2^2)^{1/2} k \tag{1.7-73}$$

in Eq. (1.3-91). It is apparent how remarkably well this approximation reproduces the exact solutions of the eigenvalue equations, particularly far from cutoff. Finally, Fig. 1.7.3 shows the results of Goell's computer analysis as dash–dotted curves. The agreement between the exact solutions of the approximate theory and the computer solutions of our problem is so close that the two types of curves are indistinguishable over most of their range. Only very close to the cutoff point is there a noticeable deviation. The discrepancy between the approximate analytical solutions and the computer solutions is most pronounced for the fundamental E_{11}^x and E_{11}^y modes. (Goell has produced very interesting pictures of the power distribution of the various modes which are shown in his paper [Gl1].)

Lack of space prevents us from discussing such interesting applications of our theory as directional couplers formed with two rectangular dielectric

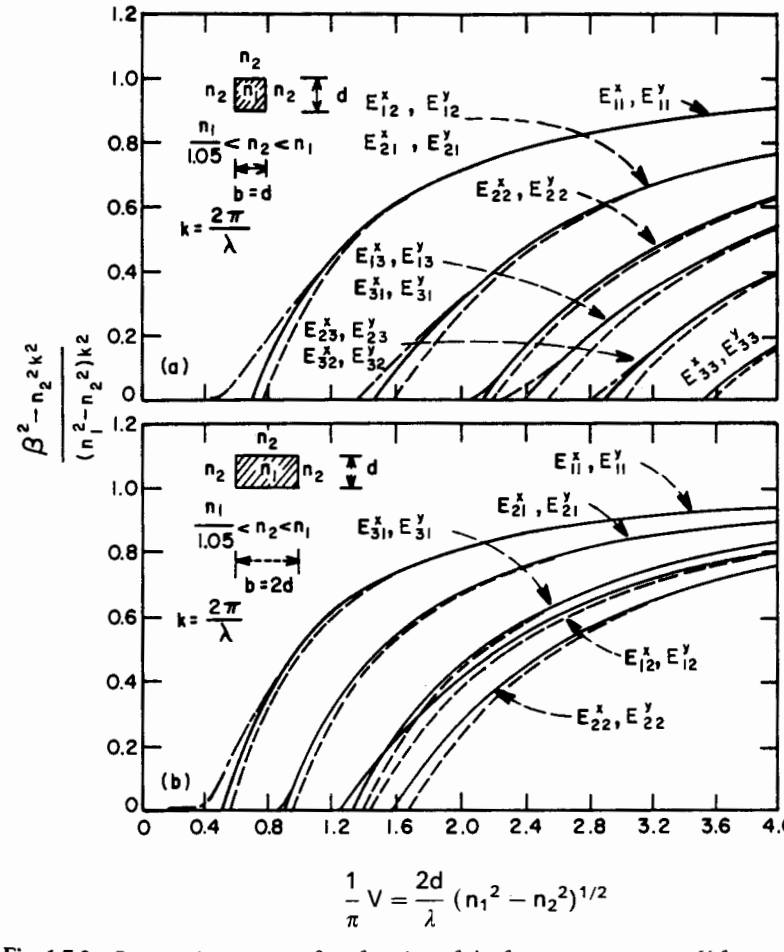

Fig. 1.7.3 *Propagation constant β as function of the frequency parameter V for two types of rectangular dielectric waveguides. The solid curves represent solutions of the eigenvalue equations of the approximate theory, the dashed curves were obtained from the far from cutoff approximations, the dash–dotted curves are Geoll's results.* (From *E. A. J. Marcatili*, "*Dielectric rectangular waveguide and directional coupler for integrated optics*," Bell. Syst. Tech. J. *48, 2071–2102 (1969). Copyright 1969, AT&T, reprinted by permission.*)

waveguides. This application is also contained in Marcatili's paper. The simplest method of computing the coupling coefficient between two adjacent dielectric waveguides consists in applying the coupling formula (10.2-18) or, in case of identical waveguides, the simplified formula (10.2-22) of [Me1] to our mode field expressions.

Chapter 2

WEAKLY GUIDING OPTICAL FIBERS

2.1 Introduction

The asymmetric slab waveguides and rectangular dielectric waveguides discussed in Chap. 1 are used in integrated optics circuits. The distances bridged by these types of waveguides are usually very short, no more than a few centimeters. We now turn our attention to optical waveguides for long distance communications. Such guides may typically be a few kilometers in length. The most promising waveguide for long distance optical communications purposes is the round optical fiber, which consists of a core of a dielectric material with refractive index n_1 and a cladding of another dielectric material whose refractive index n_2 is less than n_1. Interest in optical fibers of this type soared after it could be demonstrated that low-loss fibers are indeed feasible. The glass that is routinely used in optical instruments is far too lossy to be useful for optical fibers suitable for bridging long distances. Losses on the order of 1000 dB/km are not at all unusual and are found in high quality optical glasses used for the manufacture of lenses. Losses on the order of 100 dB/km

are exceptionally good for typical high-quality optical materials. However, the loss figures that are presently being discussed and that have been experimentally demonstrated are on the order of a few decibels per kilometer [KKM, KSZ]. Optical fibers with such low losses are very exciting and open up the possibility of new types of communications systems [KH1, MS2].

The theory of optical fibers is well understood and has been described in several books [Kyl, KB1, Me1]. However, the complete description of the guided and radiation modes of optical fibers is complicated, so that its application to the theory of fibers with imperfection is not very practical. It is our purpose to present a simplified description of the theory of optical fibers which results in simple expressions that can be used to handle problems of mode conversion and radiation loss phenomena caused by waveguide imperfections.

The exact description of the modes of round fibers is involved, since they are six component hybrid fields of great mathematical complexity. The simplification in the description of these modes is made possible by the realization that most fibers for practical applications use core materials whose refractive index is only very slightly higher than that of the surrounding cladding. By introducing the assumption that

$$n_1 - n_2 \ll 1 \tag{2.1-1}$$

considerable simplifications result. Instead of a six-component field only four field components need to be considered, and the field description is further simplified by use of rectangular cartesian instead of cylindrical coordinates. The eigenvalue equation of the simplified modes is far less complex than the corresponding equation of the exact mode fields. The simplified eigenvalue equation allows us to obtain simple approximate solutions for the regions close to cutoff and far from cutoff.

It was demonstrated in [Me1] that the simplified modes can be obtained by superposition of two nearly degenerate modes of the exact solution of the boundary value problem. This method has the advantage of establishing the relation of the approximate modes with the exact modes and showing clearly the limitations and the extent of the approximations. In this book we use the opposite approach and derive the simplified modes directly from Maxwell's equations. This procedure has the advantage of greater simplicity, but we lose the ability to judge the validity of the approximation.

Snyder [Sr1] first realized the simplification that results in the eigenvalue equation if condition (2.1-1) is applicable. Marcatili and Gloge [Ge1] perfected the description of the simplified modes of the round fiber. It was Gloge who first used the term "weakly guiding fiber" for a waveguide whose core and cladding have very nearly the same refractive index.

2.2 Guided Modes of the Optical Fiber

All dielectric waveguides support a finite number of guided modes in addition to the infinite continuum of radiation modes that are not guided by the structure but are, nevertheless, solutions of the same boundary value problem. We begin the approximate analytical treatment of round optical fibers by deriving the guided modes of the structure.

Marcatili realized that the description of the modes of the weakly guiding fiber becomes much simpler if the components of the field vectors are expressed in rectangular cartesian coordinates instead of the cylindrical coordinates that appear so much more suitable to the cylindrical geometry of the waveguide.

The cross section of the optical fiber is shown in Fig. 2.2.1. Region 1 with refractive index n_1 is the fiber core, region 2 with index n_2 is the cladding. In all our work we assume that the cladding is infinitely extended, in spite of the fact that it has a finite radius for practical fibers. The justification for assuming an infinitely extended cladding region comes from the fact that the guided modes have exponentially decaying fields outside the core. If the cladding radius is large enough, the guided mode fields have decayed to insignificant values at the outer boundary of the cladding. All practical fibers are designed to ensure that the guided mode field does not reach the outer boundary of the cladding. In the opposite case the fiber would suffer high radiation losses, since the outer fiber surface is never perfectly smooth on account of accumulating dust and other environmental effects.

The assumption of an infinite cladding radius is more questionable when we study the radiation modes. These solutions of the boundary value problem

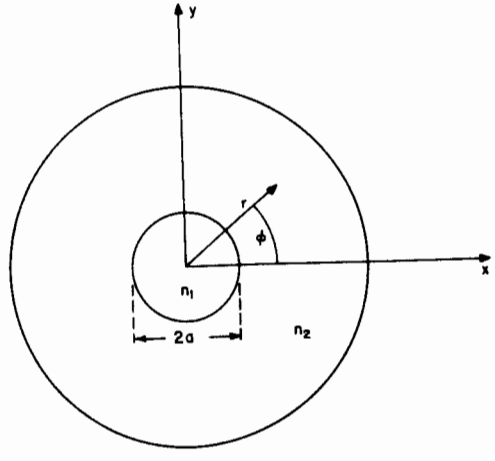

Fig. 2.2.1 *Cross section of a round optical fiber.*

reach out to infinity with undiminished strength and are certainly strongly affected by the outer cladding boundary. However, we can justify our procedure even for the radiation modes. The finite radius of the cladding has the effect of trapping some of the radiation field, causing cladding modes to appear. These modes have a discrete spectrum of allowed values of their propagation constants. But the density of these modes is much higher than the core modes, so that they form almost a continuum. When we calculate the interaction of the core modes with the radiation modes of the fiber with infinite cladding, we must keep in mind that in actuality we would have coupling of the core modes and the cladding modes. The fact that a portion of the radiation field does not actually escape freely but finds itself trapped by reflection at the outer cladding boundary does not alter our conclusions very much. In most practical cases cladding modes will be suppressed by a lossy coating on the outside of the fiber, or they will scatter out of the cladding on account of the rough outer surface. In either case, power will not endure very long in cladding modes and may be considered as being lost, just as it would be had it radiated away freely. In those cases where these conditions are not met it is necessary to study the interaction of core and cladding modes in detail.

The derivation of the simplified guided modes of the fiber uses again the longitudinal E_z and H_z components from which the transverse components are derived by means of Eqs. (1.7-4)–(1.7-7). The longitudinal components must satisfy the reduced wave Eq. (1.7-9) which we now express in cylindrical coordinates:

$$\frac{\partial^2\psi}{\partial r^2} + \frac{1}{r}\frac{\partial\psi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\phi^2} + K_j^2\psi = 0 \tag{2.2-1}$$

with

$$K_j^2 = n_j^2k^2 - \beta^2 \tag{2.2-2}$$

and

$$k = \omega(\varepsilon_0\mu_0)^{1/2} = 2\pi/\lambda \tag{2.2-3}$$

Equation (2.2-1) is obtained from Maxwell's equations by eliminating the transverse field components and solving Maxwell's equations for either E_z or H_z. All field components have the common factor

$$\exp[i(\omega t - \beta z)] \tag{2.2-4}$$

which we omit from the equations for brevity.

To solve the reduced wave equation we substitute the trial solution

$$\psi = F(r)\cos v\phi \tag{2.2-5}$$

into Eq. (2.2-1), with the result

$$\frac{d^2F}{dr^2}+\frac{1}{r}\frac{dF}{dr}+\left(K_j^2-\frac{\nu^2}{r^2}\right)F=0 \tag{2.2-6}$$

The same result would have been obtained had we used the sine instead of the cosine function in Eq. (2.2-5). Equation (2.2-6) is Bessel's differential equation. Its solutions are cylinder functions which we indicate collectively by the symbol

$$F(r)=Z_\nu(K_j r) \tag{2.2-7}$$

where $Z_\nu(K_j r)$ may be either a Bessel or a Neumann function. As solutions of a second-order differential equation there must be two independent types of functions. In the core region we need to identify $Z_\nu(K_1 r)$ with a Bessel function, $J_\nu(K_1 r)$, since the solution must remain finite at $r=0$.

When we derived the approximate solutions of the modes of the rectangular waveguide we used E_z and H_z in such a way that one of the transverse field components vanished. For the same reason we are now using the following solution of Eq. (2.2-1) in the region $r<a$:

$$E_z=\frac{iA\kappa}{2\beta_\nu}\left[J_{\nu+1}(\kappa r)\begin{Bmatrix}\sin(\nu+1)\phi\\-\cos(\nu+1)\phi\end{Bmatrix}+J_{\nu-1}(\kappa r)\begin{Bmatrix}\sin(\nu-1)\phi\\-\cos(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-8}$$

$$H_z=-\frac{iA\kappa}{2k}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}\left[J_{\nu+1}(\kappa r)\begin{Bmatrix}\cos(\nu+1)\phi\\\sin(\nu+1)\phi\end{Bmatrix}-J_{\nu-1}(\kappa r)\begin{Bmatrix}\cos(\nu-1)\phi\\\sin(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-9}$$

We have combined two solutions with $\nu+1$ and $\nu-1$ in order to achieve cancellation of the E_x and H_y components. Instead of K_1 we have used our usual notation

$$\kappa=K_1=(n_1^2k^2-\beta_\nu^2)^{1/2} \tag{2.2-10}$$

The two choices of sine and cosine functions are intended to obtain a complete set of orthogonal functions. Since both sine and cosine functions are allowed as solutions of Eq. (2.2-1), both functions are needed. The upper and lower set of functions in Eqs. (2.2-8) and (2.2-9) belong together, respectively. We now obtain the transverse components from Eqs. (2.2-8) and (2.2-9) with the help of Eqs. (1.7-4)–(1.7-7). In order to be able to carry out the derivations, we need the relations

$$\frac{\partial}{\partial x}=\frac{\partial r}{\partial x}\frac{\partial}{\partial r}+\frac{\partial\phi}{\partial x}\frac{\partial}{\partial\phi} \tag{2.2-11}$$

and

$$\frac{\partial}{\partial y}=\frac{\partial r}{\partial y}\frac{\partial}{\partial r}+\frac{\partial\phi}{\partial y}\frac{\partial}{\partial\phi} \tag{2.2-12}$$

Using the transformations

$$r=(x^2+y^2)^{1/2} \tag{2.2-13}$$

and

$$\phi=\arctan(y/x) \tag{2.2-14}$$

we obtain

$$\partial r/\partial x=x/r=\cos\phi, \tag{2.2-15}$$

$$\partial r/\partial y=y/r=\sin\phi, \tag{2.2-16}$$

$$\partial\phi/\partial x=-y/r^2=-(1/r)\sin\phi, \tag{2.2-17}$$

and

$$\partial\phi/\partial y=x/r^2=(1/r)\cos\phi \tag{2.2-18}$$

At this point we use condition (2.1-1) to simplify the analysis. Since the propagation constant of the guided mode must lie within the limits

$$n_2k<\beta<n_1k \tag{2.2-19}$$

we use approximately

$$\beta=nk \tag{2.2-20}$$

with $n\approx n_1\approx n_2$. With the help of the functional relations of the Bessel function, we obtain the following expressions for the transverse field components

$$E_y=AJ_\nu(\kappa r)\begin{Bmatrix}\cos\nu\phi\\\sin\nu\phi\end{Bmatrix} \tag{2.2-21}$$

and

$$H_x=-nA\frac{\beta_\nu}{|\beta_\nu|}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}J_\nu(\kappa r)\begin{Bmatrix}\cos\nu\phi\\\sin\nu\phi\end{Bmatrix} \tag{2.2-22}$$

The remaining components are $E_x=0$ and $H_y\approx 0$.

Outside the core in the cladding region at $r>a$, we obtain similarly

$$E_z=-\frac{A\gamma}{2\beta_\nu}\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)}\times\left[H_{\nu+1}^{(1)}(i\gamma r)\begin{Bmatrix}\sin(\nu+1)\phi\\-\cos(\nu+1)\phi\end{Bmatrix}+H_{\nu-1}^{(1)}(i\gamma r)\begin{Bmatrix}\sin(\nu-1)\phi\\-\cos(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-23}$$

$$H_z = \frac{A\gamma}{2k}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)} \times\left[H_{\nu+1}^{(1)}(i\gamma r)\begin{Bmatrix}\cos(\nu+1)\phi\\ \sin(\nu+1)\phi\end{Bmatrix} - H_{\nu-1}^{(1)}(i\gamma r)\begin{Bmatrix}\cos(\nu-1)\phi\\ \sin(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-24}$$

$$E_y = A\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)}H_\nu^{(1)}(i\gamma r)\begin{Bmatrix}\cos\nu\phi\\ \sin\nu\phi\end{Bmatrix} \tag{2.2-25}$$

$$H_x = -nA\frac{\beta_\nu}{|\beta_\nu|}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)}H_\nu^{(1)}(i\gamma r)\begin{Bmatrix}\cos\nu\phi\\ \sin\nu\phi\end{Bmatrix} \tag{2.2-26}$$

The Hankel function of the first kind, $H_\nu^{(1)}$, has now replaced the Bessel function. The choice of this function is dictated by the requirement that the field must vanish at $r \to \infty$. Only if this requirement is satisfied do we obtain a guided mode. The parameter γ appearing in the argument of the Hankel function is defined by the usual expression [compare Eq. (1.2-14)]

$$\gamma = -iK_2 = (\beta_\nu^2 - n_2^2k^2)^{1/2} \tag{2.2-27}$$

For guided modes, whose propagation constants must fall in the interval (2.2-19), γ is a real quantity. This means that the argument of the Hankel function is imaginary. The amplitude coefficient of these field expressions was chosen to ensure that E_y remains continuous at the core boundary $r = a$. Since the ratios γ/β and κ/β are both small, it is apparent that the longitudinal field components are much smaller than the transverse components. The fields are thus almost transverse. The electric field of the mode that is described by Eqs. (2.2-8), (2.2-9), and (2.2-21)–(2.2-26) is polarized in y direction. For a complete field description we also need the mode with the orthogonal polarization. For $r < a$ we obtain by the same arguments as before

$$E_z = \frac{iA\kappa}{2\beta_\nu}\left[J_{\nu+1}(\kappa r)\begin{Bmatrix}\cos(\nu+1)\phi\\ \sin(\nu+1)\phi\end{Bmatrix} - J_{\nu-1}(\kappa r)\begin{Bmatrix}\cos(\nu-1)\phi\\ \sin(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-28}$$

$$H_z = \frac{iA\kappa}{2k}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2} \times\left[J_{\nu+1}(\kappa r)\begin{Bmatrix}\sin(\nu+1)\phi\\ -\cos(\nu+1)\phi\end{Bmatrix} + J_{\nu-1}(\kappa r)\begin{Bmatrix}\sin(\nu-1)\phi\\ -\cos(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-29}$$

$$E_x = AJ_\nu(\kappa r)\begin{Bmatrix}\cos\nu\phi\\ \sin\nu\phi\end{Bmatrix} \tag{2.2-30}$$

$$H_y = nA\frac{\beta_\nu}{|\beta_\nu|}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}J_\nu(\kappa r)\begin{Bmatrix}\cos\nu\phi\\ \sin\nu\phi\end{Bmatrix} \tag{2.2-31}$$

and the field in the cladding for $r > a$ is given by

$$E_z = -\frac{A\gamma}{2\beta_\nu}\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)} \times\left[H_{\nu+1}^{(1)}(i\gamma r)\begin{Bmatrix}\cos(\nu+1)\phi\\ \sin(\nu+1)\phi\end{Bmatrix} - H_{\nu-1}^{(1)}(i\gamma r)\begin{Bmatrix}\cos(\nu-1)\phi\\ \sin(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-32}$$

$$H_z = -\frac{A\gamma}{2k}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)} \times\left[H_{\nu+1}^{(1)}(i\gamma r)\begin{Bmatrix}\sin(\nu+1)\phi\\ -\cos(\nu+1)\phi\end{Bmatrix} + H_{\nu-1}^{(1)}(i\gamma r)\begin{Bmatrix}\sin(\nu-1)\phi\\ -\cos(\nu-1)\phi\end{Bmatrix}\right] \tag{2.2-33}$$

$$E_x = A\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)}H_\nu^{(1)}(i\gamma r)\begin{Bmatrix}\cos\nu\phi\\ \sin\nu\phi\end{Bmatrix} \tag{2.2-34}$$

$$H_y = nA\frac{\beta_\nu}{|\beta_\nu|}\left(\frac{\varepsilon_0}{\mu_0}\right)^{1/2}\frac{J_\nu(\kappa a)}{H_\nu^{(1)}(i\gamma a)}H_\nu^{(1)}(i\gamma r)\begin{Bmatrix}\cos\nu\phi\\ \sin\nu\phi\end{Bmatrix} \tag{2.2-35}$$

Now we have $E_y = 0$ and $H_x \approx 0$. We have collected the field expressions for two types of guided modes whose transverse fields are polarized orthogonally to each other. These field expressions are approximate solutions of Maxwell's equations, but they do not yet satisfy all the boundary conditions requiring the tangential field components to pass continuously through the core boundary. In order to be able to satisfy the boundary conditions, we need to transform the transverse field components to cylindrical coordinates. For a general vector field we obtain the r and ϕ components from the x and y components by means of the transformations

$$F_r = F_x\cos\phi + F_y\sin\phi \tag{2.2-36}$$

and

$$F_\phi = -F_x\sin\phi + F_y\cos\phi \tag{2.2-37}$$

Since either the x or y component of the electric field vanishes, it is apparent that the E_ϕ component is simply proportional to the E_x or E_y components with the same proportionality factor (the function $\cos\phi$ or $\sin\phi$) for the fields on either side of the core boundary. Since we have adjusted the field amplitudes to ensure that the E_y or E_x components have the same value on either side of the boundary, we see that continuity of the E_ϕ components is already achieved.

The H_ϕ components, if treated exactly, would be multiplied with n_1 and n_2 inside and outside the core. However, to the approximation that ignores the difference of the two refractive indices, we see that continuity of E_ϕ results in approximate continuity of H_ϕ. We thus need to consider only the E_z and H_z components.

When we equate the E_z components on either side of the core boundary, we must equate the coefficients of $\sin(\nu+1)\phi$ [or $\cos(\nu+1)\phi$] and $\sin(\nu-1)\phi$ [or $\cos(\nu-1)\phi$] separately. Starting with the coefficients of $\sin(\nu+1)\phi$ [or $\cos(\nu+1)\phi$] we obtain from Eqs. (2.2-8) and (2.2-23) [or alternately from Eqs. (2.2-28) and (2.2-32)] the following equation:

$$\kappa J_{\nu+1}(\kappa a)/J_\nu(\kappa a) = i\gamma H^{(1)}_{\nu+1}(i\gamma a)/H^{(1)}_\nu(i\gamma a) \tag{2.2-38}$$

Exactly the same equation results also from the requirement that the H_z component be continuous at the interface.

When we equate the coefficient of $\sin(\nu-1)\phi$ [or $\cos(\nu-1)\phi$], we obtain the equation

$$\kappa J_{\nu-1}(\kappa a)/J_\nu(\kappa a) = i\gamma H^{(1)}_{\nu-1}(i\gamma a)/H^{(1)}_\nu(i\gamma a) \tag{2.2-39}$$

Equations (2.2-38) and (2.2-39) are two different forms of the eigenvalue equation of the simplified modes of the weakly guiding round fiber. If these two equations were not identical, our problem would have no solution, since we cannot have two different equations for the one unknown—the propagation constant. To show the equivalence of the two forms we use the recursion relation of the cylinder functions [JE1, GR1]

$$Z_{\nu+1}(x) = (2\nu/x)\,Z_\nu(x) - Z_{\nu-1}(x) \tag{2.2-40}$$

which holds for the Bessel and Hankel functions. When this recursion relation is used in Eq. (2.2-38) to replace the functions of order $\nu+1$ with the functions of order $\nu-1$, Eq. (2.2-39) is obtained.

The eigenvalue Eq. (2.2-39) is amazingly simple compared to the complicated form of the exact eigenvalue equation. [See, for example, Eq. (8.2-49) of [Mel].] However, simplicity has to be bought at a price. The exact eigenvalue equation has twice as many solutions as the simple Eq. (2.2-39). The exact field solutions of the round optical fiber are classified as $HE_{\nu\mu}$ or $EH_{\nu\mu}$ modes [Sn1, Mel]. The propagation constants of $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes are almost identical. They become exactly the same in the limit $n_1 \to n_2$. We are thus faced with a case of near degeneracy. Comparison of the simplified mode solutions with the exact modes shows that the simplified modes are actually a superposition of $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes [Mel]. The near degeneracy of the exact theory has thus become a definite degeneracy, and the two types of modes have merged into one. However, the total number of modes is the same in both theories, because we now have a fourfold degeneracy since both

polarizations and both choices of sine or cosine functions lead to the same eigenvalue equation.

The dispersion curves, representing the propagation constants as functions of frequency, are very nearly the same for the simplified and exact modes in case of weakly guiding fibers. Owing to the near degeneracy of the HE and EH modes, their dispersion curves are almost indistinguishable. The simplified description is thus able to reproduce the dispersion characteristics of the modes. This enables us to study the problem of pulse distortion with the use of the simplified eigenvalue equation. Problems of mode conversion and radiation losses can also be studied with the help of the simplified modes. Instead of determining how each HE or EH mode couples to other modes we now find how the superpositions of $HE_{\nu+1,\mu}$ modes and $EH_{\nu-1,\mu}$ modes couple to each other and to radiation. For purposes of determining the power transfer between groups of guided modes and for the study of radiation losses we can gain all the information that is required.

However, in spite of the obvious advantages of the simplified theory, it is prudent to keep in mind that the simplified modes do not represent true modes in the usual sense of the word. Even though we cannot determine this fact from the approximate analysis, comparison with the exact theory teaches us that the simplified modes must decompose as they travel along the waveguide. Because they are actually superpositions of $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes that travel with slightly different velocities, the simplified modes change their shape as they travel along the guide. This feature of the simplified modes becomes clear when we realize that the field shape of the superposition of two modes depends on their relative phase relationships. Because of their different phase velocities the relative phases of the $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes keep changing as a function of z, so that the superposition fields also change their shape. Only after a distance corresponding to one beat wavelength does the original phase relationship, and therefore the field shape, restore itself. The beat wavelength between modes 1 and 2 is defined as $\Lambda = 2\pi/(\beta_1 - \beta_2)$.

This discussion makes it clear that the labeling of the simplified modes does not coincide with the conventional mode labels. A simplified mode with the label ν corresponds to an $HE_{\nu+1,\mu}$ and an $EH_{\nu-1,\mu}$ mode. In particular, the important dominant HE_{11} mode is now characterized by the label $\nu = 0$. The label μ is used to distinguish the different solutions of the eigenvalue equation for a given value of ν.

The amplitude coefficient A of the guided modes can be related to the power P carried by the modes with the help of

$$P = \frac{1}{2}\int_0^{2\pi}\int_0^{\infty} (E_x H_y^* - E_y H_x^*) r\, dr\, d\phi \tag{2.2-41}$$

which follows from the power density vector Eq. (1.3-45) by integration over

the infinite cross section. We obtain the same relation for all four types of modes:

$$A = \left[\frac{4(\mu_0/\varepsilon_0)^{1/2}\gamma^2 P}{e_\nu \pi a^2 n(n_1{}^2 - n_2{}^2)k^2 |J_{\nu-1}(\kappa a)\, J_{\nu+1}(\kappa a)|}\right]^{1/2} \tag{2.2-42}$$

with

$$e_\nu = \begin{cases} 2, & \text{for } \nu = 0 \\ 1, & \text{for } \nu \neq 0 \end{cases} \tag{2.2-42a}$$

A derivation of this formula is given at the end of this section.

The eigenvalue Eq. (2.2-39) must be solved by numerical techniques. However, near cutoff and far from cutoff of each mode, approximate closed-form solutions can be worked out. Near cutoff we have the inequality

$$\gamma a \ll 1 \tag{2.2-43}$$

We can thus use the approximations of the Hankel functions for small arguments. For $\nu = 0$, we use

$$H_0^{(1)}(i\gamma a) = (2i/\pi)\ln(\Gamma\gamma a/2) \tag{2.2-44}$$

with

$$\Gamma = 1.781672 \tag{2.2-45}$$

For $\nu \neq 0$, the following approximation holds:

$$H_\nu^{(1)}(i\gamma a) = -[i(\nu-1)!/\pi](2/i\gamma a)^\nu \tag{2.2-46}$$

Substitution of these relations into Eq. (2.2-38) yields, for $\nu = 0$,

$$\gamma a = (2/\Gamma)\exp[-(1/\kappa a)\, J_0(\kappa a)/J_1(\kappa a)] \tag{2.2-47}$$

At cutoff we have $\gamma a = 0$ and $\kappa_c d = V_c$. This last relation follows from

$$(\kappa^2 + \gamma^2)a^2 = V^2 = (n_1{}^2 - n_2{}^2)k^2 a^2 \tag{2.2-48}$$

We thus obtain the cutoff condition for $\nu = 0$ modes from Eq. (2.2-47):

$$J_1(V_c) = 0 \tag{2.2-49}$$

It is apparent from Eq. (2.2-48) that near cutoff we have, to a good approximation,

$$\kappa a \approx V_c + (V - V_c) - [(\gamma a)^2/2V_c] \tag{2.2-50}$$

so that we finally find the near cutoff approximation for $\nu = 0$ modes [neglecting the $(\gamma a)^2$ term in Eq. (2.2-50)]

$$\gamma a = \frac{2}{\Gamma}\exp\left[-\frac{1}{V}\frac{J_0(V)}{J_1(V)}\right] \tag{2.2-51}$$

For the modes with $\nu = 1$ we obtain from Eqs. (2.2-39), (2.2-44), and (2.2-46),

$$\kappa a J_0(\kappa a)/J_1(\kappa a) = (\gamma a)^2 \ln(\Gamma\gamma a/2) \tag{2.2-52}$$

The cutoff condition for $\nu = 1$ modes is thus

$$J_0(V_c) = 0 \tag{2.2-53}$$

We use Eqs. (2.2-50) and (2.2-53) and expand the function $J_0(\kappa a)$ with the help of the functional relation for the derivative of Bessel functions,

$$J_\nu'(\kappa a) = -(\nu/\kappa a)J_\nu(\kappa a) + J_{\nu-1}(\kappa a) = (\nu/\kappa a)J_\nu(\kappa a) - J_{\nu+1}(\kappa a) \tag{2.2-54}$$

(where the prime indicates the derivative with respect to the whole argument) obtaining

$$J_0(\kappa a) = -[(V - V_c) - (\gamma a)^2/2V_c]J_1(V_c) \tag{2.2-55}$$

If we now replace κa with V_c on the left-hand side of Eq. (2.2-52) and substitute Eq. (2.2-55), we obtain, for modes with $\nu = 1$,

$$(\gamma a)^2 = 2V_c(V - V_c)/[1 - 2\ln(\Gamma\gamma a/2)] \tag{2.2-56}$$

This is still an implicit equation for γa, but it can easily be solved with a few iterations. V_c is obtained as the solution of Eq. (2.2-53).

To find the approximation for $\nu \geqslant 2$ we expand $J_{\nu-1}(\kappa a)$ with the help of Eqs. (2.2-50) and (2.2-54) and the cutoff condition

$$J_{\nu-1}(V_c) = 0, \qquad \text{for } \nu = 2, 3, \ldots \tag{2.2-57}$$

and obtain

$$J_{\nu-1}(\kappa a) = -[(V - V_c) - (\gamma a)^2/2V_c]J_\nu(V_c) \tag{2.2-58}$$

Equations (2.2-39) and (2.2-50) thus yield the near cutoff approximation for all modes with $\nu \geqslant 2$

$$\gamma a = \{2[(\nu-1)/\nu]V_c(V - V_c)\}^{1/2} \tag{2.2-59}$$

V_c is obtained as the solution of Eq. (2.2-57). The roots of the Bessel functions which solve Eqs. (2.2-49), (2.2-53), and (2.2-57) can be found in tables of Bessel functions [Jel, As1]. In Eq. (2.2-51) we did not expand the Bessel function $J_1(V)$ as we did in the other equations. The reason is twofold: The usual expansion of the type (2.2-55) or (2.2-58) would not work for the $V = 0$ solution of Eq. (2.2-49), since the expression $J_1(V_c)/V_c = 0/0$ is encountered in Eq. (2.2-54) for $V_c = 0$. It is, of course, easy to determine the limit of this undetermined ratio. The easiest approach would be to use the small argument approximation

$$J_\nu(V) = (1/\nu!)(V/2)^\nu \tag{2.2-60}$$

of the Bessel functions. However, the second reason for not approximating Eq. (2.2-51) any further is its wider applicability. A comparison of the approximate expression (2.2-51) with the exact solution of Eq. (2.2-39) [Me1] shows that the approximation works well for V values as large as $V = 1$. The agreement is poorer if we use approximations for the Bessel functions in Eq. (2.2-51).

It might appear as if $V_c = 0$ were also an admissible solution of Eq. (2.2-57). However, we can show with the help of the small argument approximation (2.2-60) and the small argument approximation (2.2-46) for the Hankel functions that $\kappa_c a = V_c = 0$ is not a solution of Eq. (2.2-39). The right-hand side of the equation vanishes for $\gamma \to 0$. On the left-hand side we have

$$\lim_{V_c \to 0} V_c J_{\nu-1}(V_c)/J_\nu(V_c) = 2\nu \tag{2.2-61}$$

This expression vanishes only if $\nu = 0$, showing that $V_c = 0$ is allowed only for $\nu = 0$ or for the cutoff condition (2.2-49). This shows that the lowest-order mode with $\nu = 0$ is theoretically not cutoff, since it is guided for arbitrarily small frequencies or arbitrarily thin cores. However, as a practical matter, we find with the help of Eq. (2.2-51), $\gamma a = 2.2 \times 10^{-10}$, for $V = 0.3$. In order to appreciate the meaning of this result we use the large argument approximation of the Hankel function

$$H_\nu^{(1)}(i\gamma r) = (2/\pi i\gamma r)^{1/2} \exp\{-i[\nu(\pi/2)+(\pi/4)]\} \exp(-\gamma r) \tag{2.2-62}$$

We thus see that for large γr the field, Eq. (2.2-25), outside the core is proportional to $\exp(-\gamma r)$,

$$E_y \propto \exp(-\gamma r) \tag{2.2-63}$$

The parameter γ is thus a measure for the radius r_m at which the field intensity has decayed to $1/e$ of its value at the core boundary. We can define the mode radius by the equation

$$r_m = a + \gamma^{-1} \tag{2.2-64}$$

The mode radius for the lowest-order mode, $\nu = 0$, with $V = 0.3$ is thus 4.5×10^9 times the core radius. This means, for all practical purposes, that the field is no longer guided. It seems advisable to let r_m become no more than $10a$. This value is reached for the lowest-order mode, $\nu = 0$, if $V = 0.95$.

Far from cutoff γa becomes large. With Eq. (2.2-62) the eigenvalue Eq. (2.2-39) assumes its far from cutoff approximation

$$\kappa a J_{\nu-1}(\kappa a)/J_\nu(\kappa a) = -\gamma a \tag{2.2-65}$$

The limiting value $\kappa = \kappa_\infty$, which is reached for $\gamma \to \infty$, is thus given as the root of the equation

$$J_\nu(\kappa_\infty a) = 0 \tag{2.2-66}$$

Using a method developed by Snyder [Sr1, Sr7] it was shown in [Me1] that an approximate solution of Eq. (2.2-65) is

$$\kappa a = \kappa_\infty a[1-(2\nu/V)]^{1/2\nu} \tag{2.2-67}$$

It can also be shown that the solution for $\nu = 0$ is indeed

$$\kappa a = \kappa_\infty a \exp(-1/V) \tag{2.2-68}$$

as one would expect by taking the limit of Eq. (2.2-67), for $\nu \to 0$.

Gloge [Ge1] introduced the notation $LP_{\nu\mu}$ modes for the approximate mode solutions discussed in this section. The first index, ν, corresponds to the integer that enters the circular functions, $\cos \nu\phi$ and $\sin \nu\phi$. The second index, μ, labels the roots of Eq. (2.2-39) for a given value of ν. The LP_{01} mode thus corresponds to the important dominant HE_{11} mode of the exact treatment of the fiber. Generally, we know that an $LP_{\nu\mu}$ mode is a superposition of $HE_{\nu+1,\mu}$ and $EH_{\nu-1,\mu}$ modes. The notation LP modes was chosen to suggest "linearly polarized modes."

For fibers supporting many modes, it is advantageous to be able to estimate the total number of modes that are possible for a given value of V. According to Eq. (2.2-67) κ is very nearly equal to κ_∞ for very large values of V. We thus obtain the total number of modes by estimating the number of roots of the equation

$$J_\nu(z_m) = 0 \tag{2.2-69}$$

This number is constrained by the condition

$$z_m \leqslant V \tag{2.2-70}$$

An approximate formula for the roots of Eq. (2.2-69) for large values of m is [AS1]

$$\kappa_\infty a = z_m = (\nu + 2m - \tfrac{1}{2})\pi/2 \tag{2.2-71}$$

with $m = 1, 2, 3, \ldots$. For constant values of z_m, the integer values of ν and m satisfying Eq. (2.2-71) lie along lines parallel to the dashed line depicted in Fig. 2.2.2. The largest value of $z_m = \kappa a$ is obtained from Eq. (2.2-48) with the cutoff value $\gamma = 0$. It is $z_m = V$. Neglecting the $\frac{1}{2}$ term in Eq. (2.2-71), we see that all values of ν and m that are allowed by Eq. (2.2-71) lie inside the triangle formed by the two coordinate axes and the slanted solid line passing through the points $m = V/\pi$ and $\nu = 2V/\pi$ in Fig. 2.2.2.

Each point with integer coordinate values ν and m represents one solution of Eq. (2.2-69) and, consequently, can be associated with one mode of a given polarization and ϕ dependence. It can be seen with the help of Fig. 2.2.3 that

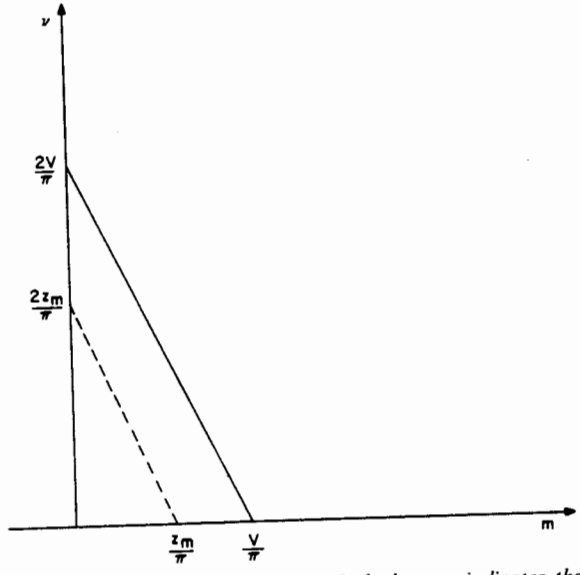

Fig. 2.2.2 *The ν, m mode number plane. The dashed curve indicates the location of the points* $\nu + 2m = \text{const}$.

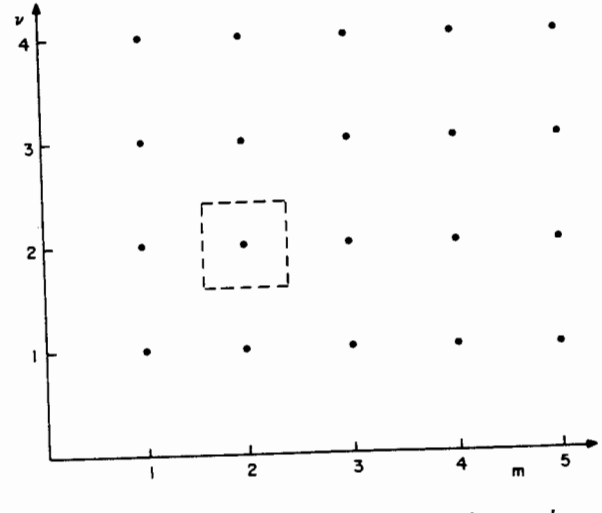

Fig. 2.2.3 *Each point represents a mode in the ν, m plane.*

each representation point occupies a square of unit area in ν, m space. The area in this space thus represents the number of modes. The total number of modes is four times the area of the triangle shown in Fig. 2.2.2, since each mode can appear in two mutually orthogonal polarizations and since we can have the sine or cosine function in Eq. (2.2-21). The area of the triangle is $(V/\pi)^2$, so that we obtain the approximate total number of modes that can exist for a given value of V from

$$N = 4V^2/\pi^2 \tag{2.2-72}$$

We found earlier [Eq. (1.3-70)] that the total number of TE and TM modes is $2V/\pi$. The total number of modes of the round fiber is thus equal to the square of the total number of slab waveguide modes for the same value of V. However, formula (2.2-72) is not very precise, since we used a number of crude approximations for its derivation. Formula (1.3-70) for the modes of the slab [with the restriction (1.3-4)] holds exactly. A better approximation of the total mode number is given by the formula [Ge1]

$$N = \tfrac{1}{2}V^2 \tag{2.2-73}$$

Finally, we study the distribution of power that is carried by the guided modes in the core and in the cladding. The amount of power that is contained inside the core is given by

$$P_{\text{core}} = \frac{1}{2}\int_0^{2\pi}\int_0^a r(E_x H_y^* - E_y H_x^*)\, dr\, d\phi \tag{2.2-74}$$

Substitution of either Eqs. (2.2-21) and (2.2-22) or (2.2-30) and (2.2-31) yields the result

$$P_{\text{core}} = e_\nu(\pi/4)\, a^2 n(\varepsilon_0/\mu_0)^{1/2} A^2 [J_\nu^2(\kappa a) - J_{\nu+1}(\kappa a) J_{\nu-1}(\kappa a)] \tag{2.2-75}$$

with

$$e_\nu = \begin{cases} 2, & \text{for } \nu = 0 \\ 1, & \text{for } \nu \neq 0 \end{cases} \tag{2.2-76}$$

For $\nu = 0$, we use the relation $J_{-1} = -J_1$. The power in the cladding is obtained similarly by integrating Eq. (2.2-74) from a to infinity:

$$\begin{aligned} P_{\text{clad}} = {} & e_\nu(\pi/4)\, a^2 n(\varepsilon_0/\mu_0)^{1/2} A^2 [J_\nu(\kappa a)/H_\nu^{(1)}(i\gamma a)]^2 \\ & \times [H_{\nu+1}^{(1)}(i\gamma a)\, H_{\nu-1}^{(1)}(i\gamma a) - H_\nu^{(1)2}(i\gamma a)] \end{aligned} \tag{2.2-77}$$

With the aid of the eigenvalue equation in its two forms (2.2-38) and (2.2-39), we can eliminate the Hankel functions from Eq. (2.2-77) and obtain

$$P_{\text{clad}} = -e_\nu(\pi/4)\, a^2 n(\varepsilon_0/\mu_0)^{1/2} A^2 [J_\nu^2(\kappa a) + (\kappa^2/\gamma^2) J_{\nu+1}(\kappa a) J_{\nu-1}(\kappa a)] \tag{2.2-78}$$

The sum of $P_{core}+P_{clad}$ is the total power P. By adding Eqs. (2.2-75) and (2.2-78), we obtain Eq. (2.2-42) with the help of Eq. (2.2-48). The negative sign in front of Eq. (2.2-78) looks disturbing, but for those γa values that are allowed by the eigenvalue equation, $J_{\nu+1}J_{\nu-1}$ is always negative, so that P_{clad} is always positive. The ratio of cladding power to total power P is obtained by substitution of Eq. (2.2-42) into Eq. (2.2-78):

$$P_{clad}/P = (1/V^2)[(\kappa a)^2+(\gamma a)^2 J_\nu^2(\kappa a)/J_{\nu-1}(\kappa a)\,J_{\nu+1}(\kappa a)] \tag{2.2-79}$$

Very far from cutoff we approximate Eq. (2.2-67):

$$\kappa a = \kappa_\infty a - \kappa_\infty a/V \tag{2.2-80}$$

Expanding the Bessel function with the help of Eqs. (2.2-80), (2.2-66), and (2.2-54) yields

$$J_\nu(\kappa a) = -\kappa_\infty a J_{\nu-1}(\kappa_\infty a)/V = \kappa_\infty a J_{\nu+1}(\kappa_\infty a)/V \tag{2.2-81}$$

and

$$J_{\nu\pm1}(\kappa a) = [1\pm(\nu\pm1)/V]\,J_{\nu\pm1}(\kappa_\infty a) \tag{2.2-82}$$

With these approximations we obtain from Eq. (2.2-79), with the help of Eqs. (2.2-48) and (2.2-80),

$$P_{clad}/P = (\kappa_\infty a)^4[1-(2/V)]/V^4 \tag{2.2-83}$$

As usual we obtain $\kappa_\infty a$ as the root of Eq. (2.2-66). The far from cutoff approximation (2.2-83) shows clearly that the amount of power carried in the cladding becomes smaller as V increases.

Near cutoff we have $\gamma a \ll 1$. For $\nu \neq 0$, we use Eq. (2.2-58) and express the ratio $J_\nu/J_{\nu+1}$ with the help of Eqs. (2.2-38) and (2.2-46) in the form

$$J_\nu(\kappa a)/J_{\nu+1}(\kappa a) = \kappa a/2\nu \tag{2.2-84}$$

Thus we can approximate Eq. (2.2-79) as follows:

$$P_{clad}/P = 1-(\gamma a)^2/2\nu[V_c(V-V_c)-\tfrac{1}{2}(\gamma a)^2] \tag{2.2-85}$$

To obtain this expression κa was replaced by V_c, which is obtained as the root of Eq. (2.2-57). The parameter γa must be obtained from Eq. (2.2-51) or (2.2-59), depending on the value of ν. For $\nu \geqslant 2$, where Eq. (2.2-59) applies, we have

$$P_{clad}/P = 1/\nu, \quad \text{for} \quad \nu \geqslant 2, \quad V = V_c \tag{2.2-86}$$

We have thus obtained the value of the power ratio at the cutoff point $V = V_c$. Actually, to this approximation, the ratio is independent of $V-V_c$. However,

because of Eq. (2.2-57) the derivative of Eq. (2.2-79) taken with respect to V is negative infinite at $V = V_c$, so that the ratio drops abruptly from the value $1/\nu$ with increasing $V-V_c$. It is interesting that P_{clad}/P does not start with the value of unity at the cutoff point of the modes with $\nu \geqslant 2$. This means that a significant amount of power resides inside the core even at the cutoff point. For the modes with $\nu = 0$ and $\nu = 1$, $P_{clad}/P = 1$ at $V = V_c$.

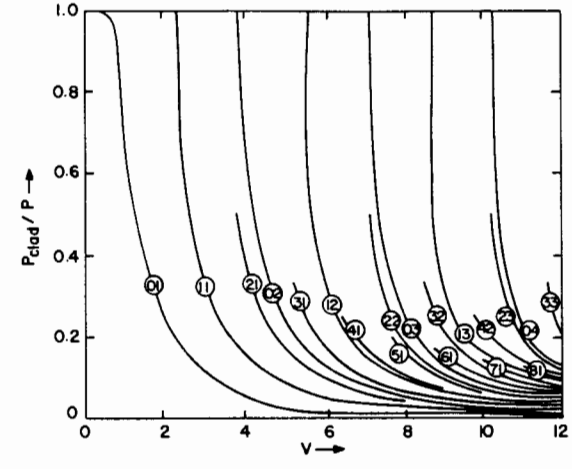

Fig. 2.2.4 *Fractional power contained in the cladding as a function of the frequency parameter V.* (From *D. Gloge*, "*Weakly guiding fibers*," Appl. Opt. *10*, *2252–2258 (1971)*, *reproduced with permission.*)

For $\nu = 0$, we have the cutoff condition (2.2-49), so that the expansion of J_1 yields, with the help of Eqs. (2.2-50) and (2.2-54),

$$J_1(\kappa a) = [V-V_c-(\gamma a)^2/2V_c]J_0(\kappa a) \tag{2.2-87}$$

Using $J_{-1} = -J_1$ and Eq. (2.2-48), we obtain from Eq. (2.2-79) near cutoff of the $\nu = 0$ modes,

$$P_{clad}/P = 1-(\gamma a)^2/[V(V-V_c)-\tfrac{1}{2}(\gamma a)^2]^2 \tag{2.2-88}$$

with γa of Eq. (2.2-47). The exponential dependence of γa ensures that the second term in Eq. (2.2-88) vanishes at $V = V_c$, even if $V_c = 0$. The power ratio is thus unity at $V = V_c$. Figure 2.2.4 shows the ratio P_{clad}/P for several modes as a function of V. The abrupt dropoff from either 1 or $1/\nu$ at the cutoff of all modes with $\nu \neq 0$ is quite apparent. Only the LP_{01} mode carries most of the power in the cladding for some distance (in terms of $V-V_c$) from its cutoff point. Figure 2.2.4 was plotted from Eq. (2.2-79) by Gloge [Ge1].

2.3 Waveguide Dispersion and Group Velocity

Having laid the theoretical groundwork for the description of the modes of the weakly guiding fiber in the preceding section we now show some numerical results.

The propagation constant β of the guided modes is obtained as a solution of the eigenvalue Eqs. (2.2-38) or (2.2-39). The propagation constant of a plane wave in a homogeneous medium with refractive index n is

$$\beta_{pw} = nk \tag{2.3-1}$$

with

$$k = \omega/c = 2\pi/\lambda \tag{2.3-2}$$

representing the plane wave propagation constant in vacuum. The difference of the refractive indices of core and cladding is only slight in weakly guiding fibers. When most of the field power is carried in the core, the wave has a propagation constant that approaches Eq. (2.3-1) with $n = n_1$. Near cutoff, most of the field energy is traveling outside the waveguide core, so that the propagation constant approaches Eq. (2.3-1) with $n = n_2$. The propagation constants of all the guided modes are thus contained in the interval

$$n_2 k < \beta < n_1 k \tag{2.3-3}$$

The approximate eigenvalue equation of weakly guiding fibers, Eq. (2.2-39), does not contain the refractive indices of core and cladding explicitly. Using relation (2.2-48) it is thus clear that its solutions, $\kappa a = f(V)$ or alternately $\gamma a = g(V)$, are universal functions of V and do not depend on the geometry and index difference of any particular waveguide. This result holds only to the approximation involved in deriving Eq. (2.2-39). By comparison, we see from Eq. (1.3-39) that the solutions for the TE modes of the symmetric slab waveguide are truly general functions of V, while the exact solutions of the eigenvalue Eq. (1.3-66) for TM modes depend on the refractive index ratios.

To the approximation applicable for weakly guiding fibers, we can define the universal function of V:

$$b = (\gamma a)^2/V^2 = [(\beta/n_2 k)^2 - 1]/[(n_1/n_2)^2 - 1] \tag{2.3-4}$$

The right-hand side of Eq. (2.3-4) is obtained from Eqs. (2.2-27) and (2.2-48). Since Eq. (2.3-4) is a universal function of V only to the extent that the weakly guiding assumption (2.1-1) applies, we can write Eq. (2.3-4) with the help of Eq. (2.3-3) in the simpler approximate form

$$b = (\beta/n_2 k - 1)/(n_1/n_2 - 1) \tag{2.3-5}$$

According to Eq. (2.3-3), b varies between 0 and 1.

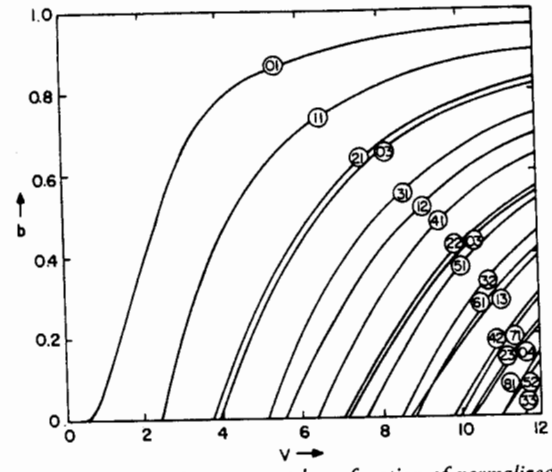

Fig. 2.3.1 *Normalized propagation constant b as function of normalized frequency V for the guided modes of the optical fiber.* (From *D. Gloge, "Weakly guiding fibers,"* Appl. Opt. **10**, *2252–2258 (1971), reproduced with permission.*)

Figure 2.3.1 shows the normalized propagation constant b as a function of V for the first 20 guided modes of the weakly guiding fiber. The number pairs written on each curve give the values of ν and μ of the $LP_{\nu\mu}$ modes. Instead of regarding b as a normalized propagation constant, we can also use the plots of Fig. 2.3.1 to obtain the important radial decay parameter γa from the relation

$$\gamma a = V\sqrt{b} \tag{2.3-6}$$

The propagation constant β is related to the phase velocity of the guided wave (ω = radian frequency) by

$$v = \omega/\beta \tag{2.3-7}$$

A more important quantity is the group velocity

$$v_g = \frac{d\omega}{d\beta} = c\frac{dk}{d\beta} \tag{2.3-8}$$

The phase velocity determines the speed at which the planes of constant phase move through the waveguide. This quantity is not directly observable at optical frequencies. The group velocity gives the speed of pulse propagation and is thus an important observable quantity. Closely related to the group velocity is the group delay. It determines the transit time of a pulse traveling through the waveguide of length L. The group delay is defined as

$$\tau_g = L\frac{d\beta}{d\omega} = \frac{L}{c}\frac{d\beta}{dk} \approx \frac{LV}{ck}\frac{d\beta}{dV} \tag{2.3-9}$$

Adopting Gloge's [Ge1] notation, we define

$$\Delta = (n_1 - n_2)/n_2 \ll 1 \tag{2.3-10}$$

and write Eq. (2.2-48) approximately as

$$V = \sqrt{2}\, n_2 \sqrt{\Delta}\, ka \tag{2.3-11}$$

Equation (2.3-9) contains the derivative of β. In order to express it in terms of the normalized propagation constant b, we obtain from Eqs. (2.3-5) and (2.3-10)

$$\beta = (b\Delta + 1)\, n_2 k \tag{2.3-12}$$

The refractive index parameter Δ is small. Since the refractive index depends only slightly on frequency, its derivative is even smaller and shall be neglected. We thus use the following approach:

$$\frac{d}{dV}(bn_2 k\Delta) = \sqrt{\Delta}\,\frac{d}{dV}(bn_2 k\sqrt{\Delta}) = \frac{\sqrt{\Delta}}{\sqrt{2}\,a}\,\frac{d(Vb)}{dV} \tag{2.3-13}$$

Using these equations we can write the total group delay as the sum

$$\tau_g = \tau_m + \tau_w \tag{2.3-14}$$

with the delay that is characteristic of the material

$$\tau_m = (L/c)\, d(n_2 k)/dk \tag{2.3-15}$$

and the waveguide delay

$$\tau_w = (L/c)\, n_2 \Delta\, d(Vb)/dV \tag{2.3-16}$$

The derivative appearing in the formula for the waveguide delay can be eliminated with the help of the eigenvalue Eq. (2.2-39). As a first step we use Eq. (2.3-4) to obtain

$$\frac{d(Vb)}{dV} = 2\frac{\gamma a}{V}\frac{d(\gamma a)}{dV} - \frac{(\gamma a)^2}{V^2} \tag{2.3-17}$$

Next we take the V derivative of the eigenvalue Eq. (2.2-39) and obtain with the help of Eqs. (2.2-40) and (2.2-54) (which also holds for the Hankel functions)

$$\kappa a\left[\frac{J_{\nu-1}(\kappa a)\, J_{\nu+1}(\kappa a)}{J_\nu^{\,2}(\kappa a)} - 1\right]\frac{d(\kappa a)}{dV} + \gamma a\left[\frac{H_{\nu-1}^{(1)}(i\gamma a)\, H_{\nu+1}^{(1)}(i\gamma a)}{[H_\nu^{(1)}(i\gamma a)]^2} - 1\right]\frac{d(\gamma a)}{dV} = 0 \tag{2.3-18}$$

We divide Eqs. (2.2-38) and (2.2-39) by γ and multiply the resulting equations, obtaining

$$\frac{H_{\nu+1}^{(1)}(i\gamma a)\, H_{\nu-1}^{(1)}(i\gamma a)}{[H_\nu^{(1)}(i\gamma a)]^2} = -\frac{(\kappa a)^2}{(\gamma a)^2}\,\frac{J_{\nu+1}(\kappa a)\, J_{\nu-1}(\kappa a)}{J_\nu^{\,2}(\kappa a)} \tag{2.3-19}$$

From Eq. (2.2-48) we derive

$$\kappa a\,\frac{d(\kappa a)}{dV} + \gamma a\,\frac{d(\gamma a)}{dV} = V \tag{2.3-20}$$

Use of these three equations and (2.3-4) enables us to eliminate the derivative from Eq. (2.3-17):

$$\frac{d(Vb)}{dV} = b\left[1 - \frac{2J_\nu^{\,2}(\kappa a)}{J_{\nu-1}(\kappa a)\, J_{\nu+1}(\kappa a)}\right] \tag{2.3-21}$$

The group delay Eq. (2.3-15), which is characteristic of the material of the optical fiber, is independent of the particular mode. The waveguide delay Eq. (2.3-16) with Eq. (2.3-21) is different for every guided mode. A light pulse that is shared by many guided modes thus splits up into many pulses arriving at the end at different times, corresponding to their group delay Eq. (2.3-16). This type of pulse dispersion is called "multimode dispersion" to distinguish it from the "material dispersion" Eq. (2.3-15).

Far from cutoff b approaches unity, while $J_\nu(\kappa a)$ goes to zero according to Eq. (2.2-66). This consideration shows that for $V \to \infty$, Eq. (2.3-21)

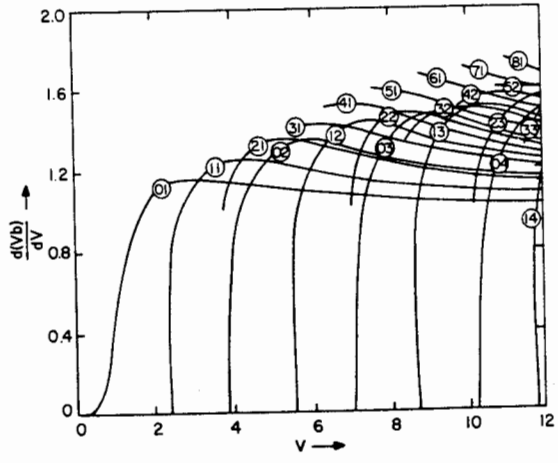

Fig. 2.3.2 *Normalized waveguide group delay as a function of V.* (From *D. Gloge,* "*Weakly guiding fibers,*" Appl. Opt. ***10***, *2252–2258 (1971), reproduced with permission.*)

approaches unity for all modes. Function (2.3-21) is plotted in Fig. 2.3.2 for the first 20 modes of the fiber. The group delay caused by waveguide dispersion is obtained from this figure with the help of Eq. (2.3-16). It is apparent from the figure that the modes with $\nu = 0$ and $\nu = 1$ have zero waveguide delay at cutoff. However, the total group delay is the sum of the material delay Eq. (2.3-15) and the waveguide delay Eq. (2.3-16), so that it always remains finite.

At cutoff $J_{\nu-1}(\kappa a) = 0$, according to Eq. (2.2-57), for all modes with $\nu \geqslant 2$. By using the approximation Eqs. (2.2-58) and (2.2-59) and expressing $J_{\nu+1}$ by J_ν with the help of Eqs. (2.2-40) and (2.2-57), we find the cutoff value of Eq. (2.3-21) for modes with $\nu \geqslant 2$,

$$d(Vb)/dV = 2(\nu-1)/\nu \tag{2.3-22}$$

The normalized propagation constant b was replaced with its value, Eq. (2.3-4), and approximated with the use of Eq. (2.2-59).

The arrival time difference between the mode with the largest waveguide group delay and the least delay [which is given by Eq. (2.3-21) assuming the value unity] is

$$\Delta\tau = (L/c)(n_1 - n_2)[1 - (2/\nu_m)] \tag{2.3-23}$$

In this formula ν_m is the largest ν value that can occur for the V at which the waveguide operates. We see from Fig. 2.2.2 that this value is approximately

$$\nu_m = 2V/\pi. \tag{2.3-24}$$

The fastest modes with $d(Vb)/dV = 0$ have been omitted from this consideration since it is unlikely that V would be close enough to the cutoff value V_c of one of these modes. Furthermore, a mode with $\nu = 0$ or $\nu = 1$ close to its cutoff point would suffer very high losses from accidental bends or irregularities of the waveguide core, so that it would not take part in the transport of energy.

Pulse distortion in multimode waveguides is caused predominantly by the different group delays of the modes. The difference $\Delta\tau$ in the arrival times of the leading and trailing edge of the pulse is given by Eqs. (2.3-23) and (2.3-24). The pulse shape depends on the amount of power that is carried by each mode. In single-mode fibers, pulse distortion is caused by the fact that the group velocity is not constant with frequency, so that different portions of the pulse spectrum travel with different velocities. This effect is much smaller than the multimode pulse distortion and depends on the frequency content of the signal and the dispersion curve of the fiber material. In addition to the material dispersion, single-mode pulse distortion is also caused by the frequency dependence of the term (2.3-16). However, for most materials the material dispersion predominates [Ge2].

2.4 Radiation Modes of the Optical Fiber

The radiation modes of the optical fiber are analogous to the radiation modes of the slab waveguide discussed in Sect. 1.4. They are a necessary supplement to the guided modes to provide a complete orthogonal set of modes. With this representation, any electromagnetic field that can exist in the presence of the fiber core can be expressed as sums over guided modes plus integrals over the various radiation modes. Just like the exact solutions for guided modes, exact solutions of radiation modes of the round optical fiber are very complicated. Since optical fibers used in most applications have very slight refractive index differences between core and cladding materials, we restrict ourselves again to approximate solutions that are valid in cases where the core–cladding refractive index difference is small.

Unlike our treatment of radiation modes of the slab waveguide, we do not consider evanescent radiation modes of the fiber; that is, we ignore modes with imaginary values of the propagation constant β that do not propagate in the z-direction. Such modes are almost never used in practical calculations and thus are not of sufficient interest to justify the added complication of having to distinguish between modes that propagate in the z-direction and those that do not. However, omitting evanescent radiation modes keeps the set of orthogonal modes from being complete.

Real optical fibers consist of cores that are surrounded by a cladding with finite radius. Radiation modes, on the other hand, are derived under the assumption that the cladding is infinitely extended. For fibers with large claddings, this may be a valid assumption and may permit us to calculate the radiation losses of a guided mode traveling in an imperfect optical fiber. Actually, the radiation field inside the cladding is constrained by the outer cladding boundary, which may be either air (vacuum) or a jacket material with a refractive index that is usually different from that of the cladding. If the refractive index of the outer jacket is smaller than that of the cladding, some of the radiation may become trapped in the cladding by total internal reflection and can be regarded as cladding modes. If the refractive index of the jacket material exceeds that of the cladding, the field is not trapped in the cladding and extends into the jacket. The behavior of cladding modes depends on the losses of the jacket material. Because of these complications, it is often assumed that the cladding is infinitely extended and radiation phenomena in the vicinity of the fiber core are described in terms of radiation modes. Whether this idealization is appropriate depends on the problem at hand and on the actual geometry of the fiber.

Assuming a small value for the refractive index difference between core and cladding materials simplifies the treatment of radiation modes in two ways. Near cutoff, that is for values of the propagation constant that are

only slightly smaller than n_2k, the radiation modes are very nearly linearly polarized, transverse waves quite similar in form to the guided modes. Far from cutoff, for values of $|\beta|$ that are substantially smaller than n_2k, the radiation modes of the fiber approach the radiation field of homogeneous space. To appreciate this phenomenon, it is helpful to remember our discussion of slab radiation modes in Sect. 1.4. There, we pointed out that radiation modes can be regarded as superpositions of radiation fields stemming from infinitely extended sources outside of the waveguide core. Waves impinging on the core at grazing angles suffer considerable reflection and refraction at the core boundary and hence are modified by its presence. Wave impinging at larger angles (measured relative to the core–cladding interface) suffer little reflection due to the slight core–cladding index difference. For these waves, it is permissible to neglect the core altogether. Thus, we design our approximate radiation modes such that they turn into the radiation field of free space (homogeneous space with constant index n_2) for small values of $|\beta|$. In the limit of vanishing core–cladding index difference, the radiation modes become exact solutions of Maxwell's equations. For values of $|\beta|$ near n_2k, the radiation modes are approximate solutions in the same sense as the weakly guided LP modes discussed in Sect. 2.2.

In Sect. 2.2, we found that there are four kinds of guided modes. The same is true for radiation modes, because for a given value of the propagation constant β there are two mutually orthogonal polarizations and for each polarization the principal electric field component can have a cosine or sine dependence on the angular coordinate ϕ. The radiation modes that most nearly resemble the y-polarized guided LP modes can be expressed as follows:

$$E_x = \frac{A}{4}\left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\sin(\nu+2)\phi\\ -\cos(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}-\sin(\nu-2)\phi\\ \cos(\nu-2)\phi\end{pmatrix}\right] \tag{2.4-1}$$

$$E_y = \frac{A}{4}\left\{2\left(\frac{nk}{|\beta|} + 1\right)Z_\nu(\rho r)\begin{pmatrix}\cos\nu\phi\\ \sin\nu\phi\end{pmatrix} - \left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\cos(\nu+2)\phi\\ \sin(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}\cos(\nu-2)\phi\\ \sin(\nu-2)\phi\end{pmatrix}\right]\right\} \tag{2.4-2}$$

$$E_z = \frac{i\rho A}{2\beta}\left[Z_{\nu+1}(\rho r)\begin{pmatrix}\sin(\nu+1)\phi\\ -\cos(\nu+1)\phi\end{pmatrix} + Z_{\nu-1}(\rho r)\begin{pmatrix}\sin(\nu-1)\phi\\ -\cos(\nu-1)\phi\end{pmatrix}\right] \tag{2.4-3}$$

$$H_x = \frac{\beta}{|\beta|}\frac{nA}{4}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left\{2\left(\frac{nk}{|\beta|} + 1\right)Z_\nu(\rho r)\begin{pmatrix}-\cos\nu\phi\\ -\sin\nu\phi\end{pmatrix} - \left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\cos(\nu+2)\phi\\ \sin(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}\cos(\nu-2)\phi\\ \sin(\nu-2)\phi\end{pmatrix}\right]\right\} \tag{2.4-4}$$

$$H_y = \frac{\beta}{|\beta|}\frac{nA}{4}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}-\sin(\nu+2)\phi\\ \cos(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}\sin(\nu-2)\phi\\ -\cos(\nu-2)\phi\end{pmatrix}\right] \tag{2.4-5}$$

$$H_z = \frac{in\rho A}{2|\beta|}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left[Z_{\nu+1}(\rho r)\begin{pmatrix}-\cos(\nu+1)\phi\\ -\sin(\nu+1)\phi\end{pmatrix} + Z_{\nu-1}(\rho r)\begin{pmatrix}\cos(\nu-1)\phi\\ \sin(\nu-1)\phi\end{pmatrix}\right] \tag{2.4-6}$$

The radiation modes corresponding to x-polarized guided LP modes have the form

$$E_x = \frac{A}{4}\left\{2\left(\frac{nk}{|\beta|} + 1\right)Z_\nu(\rho r)\begin{pmatrix}\cos\nu\phi\\ \sin\nu\phi\end{pmatrix} + \left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\cos(\nu+2)\phi\\ \sin(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}\cos(\nu-2)\phi\\ \sin(\nu-2)\phi\end{pmatrix}\right]\right\} \tag{2.4-7}$$

$$E_y = \frac{A}{4}\left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\sin(\nu+2)\phi\\ -\cos(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}-\sin(\nu-2)\phi\\ \cos(\nu-2)\phi\end{pmatrix}\right] \tag{2.4-8}$$

$$E_z = \frac{i\rho A}{2\beta}\left[Z_{\nu+1}(\rho r)\begin{pmatrix}\cos(\nu+1)\phi\\ \sin(\nu+1)\phi\end{pmatrix} - Z_{\nu-1}(\rho r)\begin{pmatrix}\cos(\nu-1)\phi\\ \sin(\nu-1)\phi\end{pmatrix}\right] \tag{2.4-9}$$

$$H_x = \frac{\beta}{|\beta|}\frac{nA}{4}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left(\frac{nk}{|\beta|} - 1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\sin(\nu+2)\phi\\ -\cos(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}-\sin(\nu-2)\phi\\ \cos(\nu-2)\phi\end{pmatrix}\right] \tag{2.4-10}$$

$$H_y = \frac{\beta}{|\beta|}\frac{nA}{4}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left\{2\left(\frac{nk}{|\beta|}+1\right)Z_\nu(\rho r)\begin{pmatrix}\cos\nu\phi\\ \sin\nu\phi\end{pmatrix} - \left(\frac{nk}{|\beta|}-1\right)\left[Z_{\nu+2}(\rho r)\begin{pmatrix}\cos(\nu+2)\phi\\ \sin(\nu+2)\phi\end{pmatrix} + Z_{\nu-2}(\rho r)\begin{pmatrix}\cos(\nu-2)\phi\\ \sin(\nu-2)\phi\end{pmatrix}\right]\right\} \tag{2.4-11}$$

$$H_z = \frac{in\rho A}{2|\beta|}\sqrt{\frac{\varepsilon_0}{\mu_0}}\left[Z_{\nu+1}(\rho r)\begin{pmatrix}\sin(\nu+1)\phi\\ -\cos(\nu+1)\phi\end{pmatrix} + Z_{\nu-1}(\rho r)\begin{pmatrix}\sin(\nu-1)\phi\\ -\cos(\nu-1)\phi\end{pmatrix}\right] \tag{2.4-12}$$

Equations (2.4-1) through (2.4-12) require detailed explanations.

The propagation constant β can assume a continuum of values in the range

$$-n_2 k < \beta < n_2 k \tag{2.4-13}$$

The propagation factor $\exp[i(\omega t - \beta z)]$, which multiplies every expression (2.4-1) through (2.4-12), has been suppressed. Continuing the practice established in Sect. 2.2 for listing the guided modes, we display the different ways of expressing the angular dependence of the field by placing the two possible choices one above the other inside a pair of parentheses. However, unlike Sect. 2.2, we now use a more compact notation by writing each field component of the radiation mode in the core and cladding regions as a single expression. Thus, the parameter ρ is defined as follows:

$$\rho \to \begin{cases} \sigma = \sqrt{n_1{}^2 k^2 - \beta^2} & \text{for } r < a \\ \rho = \sqrt{n_2{}^2 k^2 - \beta^2} & \text{for } r > a \end{cases} \tag{2.4-14}$$

The functions $Z_\nu(\rho r)$ are solutions of the differential equation (2.2-6) (with $K_j = \rho$) and assume the form

$$Z_\nu(\rho r) = \begin{cases} J_\nu(\sigma r) & \text{for } r < a \\ BH_\nu^{(1)}(\rho r) + CH_\nu^{(2)}(\rho r) & \text{for } r > a \end{cases} \tag{2.4-15}$$

If multiplied by $\sin\nu\phi$ or $\cos\nu\phi$, the Hankel functions of the first and second kind, distinguished by the superscripts 1 and 2, are cylindrical traveling waves that travel in the positive (superscript 2) and negative (superscript 1) r-direction. The direction of wave travel associated with a given Hankel function is contingent on our practice of writing the time dependence as $\exp(i\omega t)$. Had we used instead $\exp(-i\omega t)$, the waves associated with each Hankel function would change direction. In the same way, the Bessel function $J_\nu(\sigma r)$ represents a cylindrical standing wave. The refractive index n appearing in Eqs. (2.4-1) through (2.4-12) should assume the value n_1 inside the core and n_2 inside the cladding. However, since the difference

$n_1 - n_2$ is assumed to be small, we do not distinguish between these two possibilities whenever n appears in the equations without a subscript. But if n is identified as the refractive index of the prospective regions, Eqs. (2.4-1) through (2.4-12) represent exact solutions of Maxwell's equations, which, however, does not mean that they are exact solutions of the fiber problem.

As radiation modes of the optical fiber, the fields must also satisfy the boundary conditions, which, in their exact formulations, would consist of requiring that the field components, E_ϕ, E_z, H_ϕ, and H_z must remain continuous at the core boundary at $r = a$. But with the two undetermined coefficients at our disposal the four boundary conditions cannot be satisfied exactly. Instead, we require that $Z_\nu(\rho r)$ and $dZ_\nu(\rho r)/dr$ are continuous at $r = a$. These conditions are satisfied if we assign B and C the following values:

$$B = i\frac{\pi a}{4}\left[\sigma J_{\nu+1}(\sigma a)H_\nu^{(2)}(\rho a) - \rho J_\nu(\sigma a)H_{\nu+1}^{(2)}(\rho a)\right] \tag{2.4-16}$$

and

$$C = -i\frac{\pi a}{4}\left[\sigma J_{\nu+1}(\sigma a)H_\nu^{(1)}(\rho a) - \rho J_\nu(\sigma a)H_{\nu+1}^{(1)}(\rho a)\right] \tag{2.4-17}$$

In deriving these expressions from the continuity requirements for Z_ν and its r-derivative, we have used the following well-known functional relationship (Wronskian) of the Hankel functions [AS1].

$$H_{\nu+1}^{(1)}(x)H_\nu^{(2)}(x) - H_\nu^{(1)}(x)H_{\nu+1}^{(2)}(x) = \frac{4}{i\pi x} \tag{2.4-18}$$

By expressing the Hankel functions in terms of the Bessel function J_ν and the Neumann function N_ν, it is easy to show that the coefficients become $B = C = \frac{1}{2}$ in the limit $\sigma = \rho$. In that case, we have $Z_\nu(\rho r) = J_\nu(\sigma r)$ so that the two forms of Z_ν shown in (2.4-15) become identical. The proof of this assertion requires the use of another Wronskian [AS1]

$$J_{\nu+1}(x)N_\nu(x) - J_\nu(x)N_{\nu+1}(x) = \frac{2}{\pi x} \tag{2.4-19}$$

In using the expressions (2.4-15), (2.4-16), and (2.4-17), it is important to remember that B and C are constants for a given mode with azimuthal mode label ν. This means that when we form expressions such as $Z_{\nu+2}(\rho r)$ or $Z_{\nu-2}(\rho r)$ we do not also change the subscripts of the Bessel and Hankel functions appearing in (2.4-16) and (2.4-17). These expressions remain the same for a given mode. However, the subscripts of the r-dependent functions $H_\nu^{(1)}(\rho r)$ and $H_\nu^{(2)}(\rho r)$ change as the subscript of Z_ν is changed. For this reason, we refrained from attaching the ν-label to B and C since such a subscript

notation would lead to the temptation of changing the subscripts inside the brackets in (2.4-16) and (2.4-17) as the subscript of $Z_\nu(\rho r)$ is changed. The subscripts in the definitions of B and C change only when we consider another mode whose ν-label is actually different.

The expressions (2.4-1) through (2.4-12) describe the radiation modes in the entire β-range indicated by (2.4-13). However, for $|\beta|$-values near n_2k, they resemble the guided mode field inside the core. To see this, we express the difference $nk - |\beta|$ as follows:

$$n_1k - |\beta| = \frac{\sigma^2}{n_1k + |\beta|} \tag{2.4-20}$$

or

$$n_2k - |\beta| = \frac{\rho^2}{n_2k + |\beta|} \tag{2.4-21}$$

Thus, if we regard σ and ρ as small quantities, the difference $nk - |\beta|$ is small of second order. This means that for radiation modes that are just below the cutoff of guided modes, the E_x-component in (2.4-1) is a second-order quantity relative to the much larger E_y-component. The longitudinal component E_z is small to first order. Neglecting second-order quantities, we see that the field components (2.4-1) through (2.4-6) of the radiation modes are very similar in form to the corresponding field components of the guided modes listed in (2.2-8), (2.2-9), (2.2-21), and (2.2-22) (a corresponding statement is true for the radiation modes described by (2.4-7) through (2.4-12). The smallness of the E_x-component for $|\beta| \approx nk$ justifies our assertion that the fields (2.4-1) through (2.4-6) resemble the y-polarized field of the guided modes. Similarly, the fields (2.4-7) through (2.4-12) can be regarded as x-polarized near cutoff.

The boundary conditions requiring continuity of Z_ν and dZ_ν/dr at $r = a$ ensure approximate continuity of E_y and E_z for β near n_2k. The continuity of E_y of (2.4-2) follows directly from the continuity of Z_ν if we neglect the term that is multiplied by $nk - |\beta|$. To see that E_z is also continuous requires the use of the recursion relation for the derivatives of the cylinder functions

$$xZ_{\nu\pm1}(x) = \nu Z_\nu(x) \mp \frac{dZ_\nu(x)}{dx} \tag{2.4-22}$$

which can also be written in the form

$$\rho Z_{\nu\pm1}(\rho r) = \frac{\nu}{r} Z_\nu(\rho r) \mp \frac{dZ_\nu(\rho r)}{dr} \tag{2.4-23}$$

The left-hand side of (2.4-23) is contained in E_z. Thus, since Z_ν (and hence $(\nu/r)Z_\nu$) and dZ_ν/dr are continuous, E_z is continuous at $r = a$. The same argument holds for H_z. Thus, for $|\beta| \approx nk$, the differences of the core and cladding fields taken at $r = a$ contain only terms that are small to second order and hence are negligible in the spirit of our approximation. This means that the boundary conditions are satisfied for $|\beta| \approx nk$.

For values of $|\beta|$ that are smaller than n_2k, so that $nk - |\beta|$ can no longer be regarded as small, the difference $\sigma - \rho$ becomes small. According to their definitions (2.4-14), we have

$$\sigma = \sqrt{\rho^2 + (n_1^2 - n_2^2)k^2} \approx \rho\left[1 + (n_1^2 - n_2^2)\frac{k^2}{2\rho^2}\right] \tag{2.4-24}$$

The extreme right-hand side of this equation is valid as ρ becomes larger with decreasing values of $|\beta|$. Because $n_1 - n_2 \ll 1$, the distinction between σ and ρ becomes negligible for smaller values of $|\beta|$. As pointed out above, with σ approaching ρ, B and C become equal to $\frac{1}{2}$ and $Z_\nu(\rho r)$ approaches $J_\nu(\sigma r)$. In this case, the distinction between core and cladding fields vanishes so that the continuity of the fields at the core–cladding boundary remains approximately valid. Thus, the fields (2.4-1) through (2.4-12) do not only satisfy Maxwell's equations, they also satisfy the continuity conditions at $r = a$ for all values of β—at least approximately.

It can also be shown that the radiation modes are approximately orthogonal to each other to the same approximation as the boundary conditions. If we use the subscripts p and q to distinguish radiation modes that differ in their functional dependence on ϕ or in their polarization, the following orthogonality relation is approximately valid

$$\frac{1}{2}\int_0^{2\pi} d\phi \int_0^{\infty} r[\mathbf{E}_{\nu p}(r, \phi, \rho) \times \mathbf{H}^*_{\mu q}(r, \phi, \rho')]_z\, dr = \frac{\beta'}{|\beta|} P\delta_{\nu\mu}\delta_{pq}\delta(\rho - \rho') \tag{2.4-25}$$

where $\delta_{\nu\mu}$ and δ_{pq} are the Kronecker delta symbols that are unity for equal subscripts and zero for different subscripts, and $\delta(\rho - \rho')$ is the Dirac delta function. Equation (2.4-25) implies normalization with respect to P, which leads to the following expression for the amplitude coefficient A appearing in the field expressions (2.4-1) through (2.4-12)

$$A^2 = \frac{\rho|\beta|(\mu_0/\varepsilon_0)^{1/2}P}{2e_\nu \pi n^2 kBC} \tag{2.4-26}$$

with

$$e_\nu = \begin{cases} 2 & \text{for } \nu = 0 \\ 1 & \text{for } \nu \neq 0 \end{cases} \tag{2.4-27}$$

and with ρ defined by the second line on the right-hand side of (2.4-14). B and C are defined by (2.4-16) and (2.4-17).

We can use the radiation modes of the weakly guided optical fiber to recover the guided modes. Since for guided modes $|\beta| > n_2 k$, ρ of (2.4-14) becomes imaginary so that we write

$$\rho = i\gamma \tag{2.4-28}$$

For a large imaginary argument, the Hankel function $H_\nu^{(1)}(i\gamma r)$ decreases exponentially with increasing values of r. However, the Hankel function of the second kind $H_\nu^{(2)}(i\gamma r)$ grows exponentially and becomes infinitely large for $r \to \infty$. To obtain a guided mode whose field intensity vanishes at $r \to \infty$, we must demand that $C = 0$. A comparison of (2.4-17) and (2.2-38) shows that this condition represents the eigenvalue equation of weakly guided fiber modes with σ renamed as κ.

It might appear as though we have now obtained a more accurate description of the LP modes of the fiber since the field expressions (2.4-1) through (2.4-12) contain more terms than those listed in Sect. 2.2. However, it is important to realize that the additional terms are small of second order for $n_2 k < |\beta| < n_1 k$. They are thus neglible in the spirit of the weak guidance approximation and do not result in an improved representation of guided LP modes. In fact, it can be shown that these second-order terms are wrong if we try to regard them as higher approximations toward the exact guided modes of the fiber. But the additional terms in (2.4-1) through (2.4-12) are needed to represent the radiation modes in regions where $|\beta|$ is smaller than $n_2 k$ and where these terms are no longer small. They are correct and necessary parts of the radiation modes of the weakly guiding fiber.

The expressions for the radiation modes listed in (2.4-1) through (2.4-12) reveal a feature that may appear disturbing at first sight. That is, all field components become infinitely large when the propagation constant goes to zero. The pole at $\beta = 0$ is proportional to $1/\sqrt{|\beta|}$ since a factor $\sqrt{|\beta|}$ is contributed to the field expressions through the normalization according to (2.4-26). Physically, radiation modes with $\beta = 0$ represent standing waves whose incoming and outgoing components travel at right angles to the fiber axis.

This seeming paradox is resolved when we realize that a single radiation mode has no physical meaning. Radiation modes always are used in integral representations of a more general solution of a radiation problem. Thus, they always appear under an integral with respect to the variable ρ. This fact is also utilized when we normalize the radiation modes with respect to the delta function which becomes infinite for vanishing values of its argument $\rho - \rho'$.

To see that the pole of the radiation modes at $\beta = 0$ does not cause the integral to diverge, we transform the integration variable ρ to the new integration variable β. Using (2.4-14) we obtain

$$d\rho = -\frac{\beta}{\rho}\, d\beta \tag{2.4-29}$$

Thus, if we change integration variables from ρ to β, we gain an additional factor β so that now the integrand vanishes at $\beta = 0$. Thus, even though the radiation modes become infinitely large at $\beta = 0$, their contribution to any field expansion remains finite.

2.5 Cutoff and Total Internal Reflection

In Sect. 1.3 we discussed the cutoff condition for modes of the slab waveguide and pointed out its connection with the critical angle of total internal reflection. In fact, the cutoff condition for slab waveguide modes can be derived from the critical angle at which total internal reflection is lost at a plane dielectric interface. The mode fields in round optical fibers can locally be expressed as quasi-plane waves. Extending our knowledge of the cutoff condition in slab waveguides, we may be tempted to derive the cutoff condition for fiber modes by applying the idea of the critical angle of total internal reflection to the quasi-plane waves of the fiber modes. However, this procedure leads to the wrong results. For most modes, cutoff in optical fibers is not related to the loss of total internal reflection. Instead we show in this section that cutoff in optical fibers is related to the phenomenon of radiation in bent dielectric waveguides. The two explanations of cutoff—loss of total internal reflection in slabs and radiation caused by bent dielectric interfaces—do not contradict each other. The critical angle of total internal reflection is a property of plane dielectric interfaces. It is not surprising that other phenomena take over if the plane interface is being curved. The feeling of surprise is associated with our habit of thinking in terms of geometrical optics, of which total internal reflection and its critical angle are integral parts.

We base our discussion on the form (2.2-21) of the guided mode field inside the fiber core. The Bessel function is expressed by its approximation for large argument and any order number $\nu < \kappa r$ [GR1]:

$$J_\nu(\kappa r) \approx \frac{(e^{i\psi} + e^{-i\psi})}{(2\pi)^{1/2}[(\kappa r)^2 - \nu^2]^{1/4}} \tag{2.5-1}$$

with

$$\psi = [(\kappa r)^2 - \nu^2]^{1/2} - \nu \arccos(\nu/\kappa r) - (\pi/4) \tag{2.5-1a}$$

We express the argument ψ of the exponential functions by the first two terms of a Taylor series expansion in the vicinity of the point $r = a$:

$$\psi(r) \approx \psi(a) + [\kappa^2 - (\nu/a)^2]^{1/2} r + \text{const.} \tag{2.5-2}$$

Reinstating the omitted factor (2.2-4), we approximate Eq. (2.2-21) with the help of Eq. (2.5-1):

$$\begin{aligned} E_y &= A J_\nu(\kappa r) \cos \nu\phi \, e^{i(\omega t - \beta z)} \\ &\approx \frac{1}{2} \frac{(e^{i\psi} + e^{-i\psi})(e^{i\nu\phi} + e^{-i\nu\phi}) e^{-i\beta z} e^{i\omega t}}{(2\pi)^{1/2} [(\kappa r)^2 - \nu^2]^{1/4}} \end{aligned} \tag{2.5-3}$$

This approximation shows that the mode field consists of the superposition of four plane waves. Instead of traveling parallel to the waveguide axis, these quasi-plane waves spiral around the axis. The rays associated with these waves are called skew rays. The four quasi-plane waves belong to two skew rays following the path of a right- and left-handed screw. Each ray, in turn, is decomposed into a component approaching and receding from the core boundary. For our purposes it is sufficient to consider one of the two skew rays approaching the dielectric interface. The corresponding quasi-plane wave follows from Eqs. (2.5-2) and (2.5-3):

$$\exp\{-i\mathbf{K}_c \cdot \mathbf{r}\} = \exp\{-i\{[\kappa^2 - (\nu/a)^2]^{1/2} r + (\nu/a)(a\phi) + \beta z\}\} \tag{2.5-4}$$

With the unit vectors $\mathbf{e}_r$, $\mathbf{e}_\phi$, and $\mathbf{e}_z$ in r, ϕ, and z directions, we can express the vector $\mathbf{r}$ as

$$\mathbf{r} = r\mathbf{e}_r + (a\phi)\,\mathbf{e}_\phi + z\mathbf{e}_z \tag{2.5-5}$$

and write the propagation vector of the quasi-plane wave in the form

$$\mathbf{K}_c = [\kappa^2 - (\nu/a)^2]^{1/2} \mathbf{e}_r + (\nu/a)\,\mathbf{e}_\phi + \beta \mathbf{e}_z \tag{2.5-6}$$

The cosine of the angle between the propagation vector $\mathbf{K}_c$ and the direction normal to the core boundary is given by [Eq. (2.2-10) is used to simplify the denominator]

$$\cos\alpha = (1/|\mathbf{K}_c|)\,\mathbf{K}_c \cdot \mathbf{e}_r = [\kappa^2 - (\nu/a)^2]^{1/2}/n_1 k \tag{2.5-7}$$

The sine of this angle is correspondingly

$$\sin\alpha = (1 - \cos^2\alpha)^{1/2} = [(n_1 k)^2 - \kappa^2 + (\nu/a)^2]^{1/2}/n_1 k \tag{2.5-7a}$$

With the help of Eq. (2.2-10) we obtain

$$n_1 \sin\alpha = [\beta^2 + (\nu/a)^2]^{1/2}/k \tag{2.5-8}$$

The cutoff condition for guided modes of the fiber is

$$\gamma = 0 \tag{2.5-9}$$

with γ defined by Eq. (2.2-27). Equation (2.5-9) leads to

$$\beta = n_2 k \tag{2.5-10}$$

That Eq. (2.5-9) or (2.5-10) is indeed the proper cutoff condition for guided fiber modes follows from the form of the mode fields outside the fiber core. Equations (2.2-23)–(2.2-26) show that the field outside the core is expressed by Hankel functions with the argument $i\gamma r$. For imaginary argument values, the Hankel function decays exponentially at infinite distance from the fiber core. However, as γ becomes imaginary the argument of the Hankel function becomes real. It now describes a radiating cylindrical wave at infinite distances from the fiber core. The mode is no longer guided when it loses power by radiation. It is thus clear that the mode description forces us to accept Eq. (2.5-9) or (2.5-10) as the cutoff condition.

Let us now look at the angle α right at cutoff. Substitution of Eq. (2.5-10) into Eq. (2.5-8) results in

$$n_1 \sin\alpha = n_2 [1 + (\nu/n_2 ka)^2]^{1/2} \tag{2.5-11}$$

Equation (2.5-11) leads to the inequality

$$n_1 \sin\alpha > n_2, \qquad \text{for} \quad \nu \neq 0 \tag{2.5-12}$$

By definition, the critical angle for total internal reflection follows from Eq. (1.2-1), with $\alpha_2 = 0$,

$$n_1 \sin\alpha_{1c} = n_2 \tag{2.5-13}$$

We are thus forced to the conclusion that, right at the cutoff point of the guided mode, its direction of propagation with respect to the nomal to the core–cladding interface obeys the inequality

$$\alpha > \alpha_{1c} \tag{2.5-14}$$

At the cutoff point of the mode field, the rays (associated with this field) impinge on the dielectric interface with an angle that is larger than the critical angle, so that from the point of view of ray optics total internal reflection is still taking place. Only if $\nu = 0$ do we have $\alpha = \alpha_{1c}$, so that cutoff coincides with the loss of total internal reflection.

In order to understand the physical reason for the cutoff in fibers, we take a look at curved slab waveguides. It is well known that lossless transmission through curved slab waveguides is impossible [Mel, Mi2]. Consider a

guided mode traveling in a dielectric slab that is bent into a large hollow cylinder. The wave is traveling in a direction perpendicular to the axis of the cylinder. If the radius of curvature is large, the field near the curved slab is very similar to the field in the straight slab. However, far from the slab, in the direction away from the cylinder, the exponential field decay is changing into a radiation field. Since a guided mode is defined as a wave whose field tends to zero at infinity and does not lose power by radiation, we see that the wave in the curved slab waveguide is no longer a guided mode in the strict sense. In fact, in accordance with the discussion in Sect. 1.5 we may call it a leaky wave. Leaving the outer radius of the cylinder fixed but shrinking the inner radius results in a solid dielectric rod in the limit of vanishing inner radius. The slab waveguide mode is being transformed by this transition into a field configuration of the optical fiber. However, this field does not become a guided mode of the fiber but belongs to the leaky wave solutions. It is a wave beyond cutoff. The rays associated with this field impinge on the dielectric interface safely above the critical angle for total internal reflection. The loss of power is caused by the evanescent field outside the waveguide core. The wave traveling around the circumference of the cylinder drags its evanescent field along. At increasing distances from the fiber the field is forced to move faster. At a certain critical radius it would exceed the velocity of light in the medium outside the fiber core. At this point the field detaches itself and radiates away. This picture would suggest that any wave with $\nu \neq 0$ would suffer the same fate, since all waves of this type have a nonvanishing component of the propagation vector (2.5-6) in circumferential direction. In order to understand why some modes lose power by radiation while others are properly guided, we consider the field of the modes outside the fiber core. The Bessel function is now replaced by a Hankel function. Using the large argument approximation (2.2-62), we have the following expression for the field (2.2-25) outside the fiber core.

$$\begin{aligned} E_y &= BH_\nu^{(1)}(i\gamma r)\cos\nu\phi\, e^{i(\omega t-\beta z)} \\ &\approx \tfrac{1}{2}(2/\pi i\gamma r)^{1/2} Be^{i\psi}e^{-\gamma r}\left[e^{i(\nu\phi-\beta z)}+e^{-i(\nu\phi+\beta z)}\right]e^{i\omega t} \end{aligned} \tag{2.5-15}$$

The phase angle ψ assumes the constant value $\psi = (\nu+1/2)(\pi/2)$. For guided modes γ is real, so that there are now only two quasi-plane waves spiraling like right- and left-handed screws around the fiber. The propagation vector of one of these plane waves is given by

$$\mathbf{K} = (\nu/r)\mathbf{e}_\phi + \beta\mathbf{e}_z \tag{2.5-16}$$

This vector does not have a radial component. Its circumferential component decreases with increasing distance from the fiber axis. At $r = \infty$ the quasi-plane waves move parallel to the waveguide axis. The mode field outside the

fiber core has the tendency to lose its circumferential component of travel and to follow the axis of the fiber at large distances. Whether the field loses power by radiation or whether it can be a properly guided wave depends on the rate at which the circumferential velocity component decreases. We use the following expression for the velocity of the quasi-plane wave associated with the evanescent field outside the fiber:

$$v(r) = \omega/|\mathbf{K}| = kc/[(\nu/r)^2+\beta^2]^{1/2} \tag{2.5-17}$$

Equation (2.5-17) represents the phase velocity (not the group velocity) of the quasi-plane wave. However, the explanation of the loss mechanism of bent dielectric waveguide fields also involves the phase velocity.

Since we have for guided waves

$$\beta > n_2 k \tag{2.5-18}$$

we see that for all radii

$$v(r) < c/n_2 \tag{2.5-19}$$

This shows that guided waves manage to maintain the velocity of their fields outside the core below the velocity of light, c/n_2. If β becomes too small, the velocity of light is exceeded at some radius that depends on the value of ν. In particular, the wave (discussed above) with $\beta = 0$ travels in circumferential direction without an axial velocity component, so that from Eq. (2.5-17) we have

$$v(r) = rkc/\nu, \qquad \text{for} \quad \beta = 0 \tag{2.5-20}$$

At $r = \nu/(n_2 k)$ this wave exceeds the velocity of light and loses power by radiation. By maintaining inequality (2.5-18) the guided mode fields do not reach quasi-plane wave velocities in excess of the light velocity outside the waveguide core. However, Eq. (2.5-17) shows clearly that all guided modes reach the conditions

$$v(r) = c/n_2 \tag{2.5-21}$$

at $r = \infty$ at their cutoff points $\beta = n_2 k$. It is thus clear that cutoff means that the guided mode field is just beginning to move at the velocity of light at $r = \infty$. At cutoff the guided mode field detaches itself at infinite distances and radiates away. This discussion shows that even beyond cutoff the mode losses may not be very high in very thick optical fibers, since bent slab waveguides are known to have low radiation losses if their bending radius is large.

For modes with $\nu = 0$, cutoff can be explained in the same way as for slab waveguide modes. Expressions (2.5-6) and (2.5-16) for the propagation constants of the quasi-plane waves show that there are no circumferential

components of wave propagation if $v = 0$. The field is thus not moving around the axis of the guide. The cutoff condition for $v = 0$ modes follows from the critical angle argument. However, we can explain the critical angle of total internal reflection also in terms of light velocity. A guided mode field in a slab (or in a fiber with $v = 0$) is accompanied by an evanescent wave outside the core region. As cutoff is approached the velocity of the evanescent field is increasing. The direction of the quasi-plane wave propagation vectors outside the core is parallel to the interface. No power is radiating away. However, as the critical angle (or cutoff) is reached, the evanescent field is just reaching the velocity of light in the outside medium. If the direction of the quasi-plane wave propagation vector would remain parallel to the surface, the phase velocity of the field would exceed the light velocity c/n_2. This "catastrophe" is avoided by a change of the direction of the propagation vector of the quasi-plane waves. They now acquire a component perpendicular to the interface so that power is radiated away from the surface as the mode goes beyond cutoff. In all cases it is the onset of power loss by radiation that determines cutoff. The cutoff phenomenon that is related to the critical angle of total internal reflection and the cutoff for fiber modes with $v \neq 0$ are distinguished by the location where the quasiplane waves would exceed the velocity of light if no radiation would occur. For slab waveguide modes and fiber modes with $v = 0$, there is no critical distance. Below and right at cutoff, the evanescent field is moving with uniform velocity everywhere. For modes with $v \neq 0$ the critical distance is located at $r = \infty$.

Chapter 3

COUPLED MODE THEORY

3.1 Introduction

The guided modes that we studied in the preceding chapters are waveforms that can actually be excited. These waves propagate along the axis of the dielectric waveguide undisturbed provided that the waveguide structure is free of imperfections. If the dielectric materials of the guide are lossy, the propagation constant β becomes complex. However, for reasonable low loss the solutions presented in the last two chapters are still valid. The losses of the modes of weakly guiding fibers are equal to the plane wave loss in the dielectric material of the guide provided that core and cladding have the same losses.

Actual waveguides are never perfect. There are always index inhomogeneities or slight changes of the core width. These imperfections cause the modes of the waveguide to couple among each other. If we excite a pure mode at the beginning of the guide, some of its power is transferred to other guided modes and also to the radiation modes. Power transfer to other guided modes results in signal distortion since each guided mode travels at its own characteristic