A waveguide is an electromagnetic feed line used in microwave communications, broadcasting, and radar installations. A waveguide consists of a rectangular or cylindrical metal tube or pipe. The electromagnetic field propagates lengthwise. Waveguides are most often used with horn antenna s and dish antenna s.

An electromagnetic field can propagate along a waveguide in various ways. Two common modes are known as transverse-magnetic (TM) and transverse-electric (TE). In TM mode, the magnetic lines of flux are perpendicular to the axis of the waveguide. In TE mode, the electric lines of flux are perpendicular to the axis of the waveguide. Either mode can provide low loss and high efficiency as long as the interior of the waveguide is kept clean and dry.

To function properly, a waveguide must have a certain minimum diameter relative to the wavelength of the signal. If the waveguide is too narrow or the frequency is too low (the wavelength is too long), the electromagnetic fields cannot propagate. At any frequency above the cutoff (the lowest frequency at which the waveguide is large enough), the feed line will work well, although certain operating characteristics vary depending on the number of wavelengths in the cross section.

Abstract. In a planar regular optical waveguide, propagation of polarized monochromatic electromagnetic radiation obeys a law following from the Maxwell equations. The Maxwell equations in Cartesian coordinates associated with the waveguide geometry can be written as the two independent systems of equations:

$$E_x = \frac{\beta}{\varepsilon} H_y,\ \frac{dE_z}{dx} = \frac{ik_0}{\varepsilon}\left(\varepsilon\mu - \beta^2\right) H_y,\ \frac{dH_y}{dx} = ik_0 \varepsilon E_z,$$

$$H_x = -\frac{\beta}{\mu} E_y,\ \frac{dH_z}{dx} = -\frac{ik_0}{\mu}\left(\varepsilon\mu - \beta^2\right) E_y,\ \frac{dE_y}{dx} = -ik_0 \mu H_z.$$

Each of the systems can be transformed to a second order ODE for the leading component and two other equations for straightforward computation of the complementary electromagnetic field components. In doing so, the boundary conditions for Maxwell's equations are reduced to two pairs of boundary conditions for obtained equations. In addition, the asymptotic conditions hold for each class of waveguide modes. Thus, the problem of description of a complete set of modes in a regular planar waveguide is formulated in terms of the eigenvalues problem for the essentially self-adjoint second order differential operator:

$$-\frac{d^2\psi}{dx^2} + V(x)\,\psi = k^2\psi.$$

For the operator, we find some results about its spectrum, complete sets of solutions, and diagonalization by an isometric isomorphism (generalized Fourier transformation); new basis functions are related to initial ones by simple transformation formulas. The eigenvalues problem is equivalently reduced to the two problems (left and right) of the one-dimensional potential scattering theory by projection on the two branches of the continuous spectrum.

Keywords: waveguide propagation of electromagnetic radiation, equations of waveguide modes of regular waveguide, guided modes, radiation modes, a complete set of modes of a planar waveguide.

MSC numbers: 65Fxx, 65Hxx, 65L10, 65L15, 78A40, 78Mxx

1. Introduction

To describe propagation of electromagnetic radiation in integrated optical waveguides by the coupled-wave method [1, 2], by the comparison-of-waveguides method [3, 4], or by the incomplete Galerkin method [5, 6], we need to know a complete system of waveguide modes of a regular planar waveguide [7, 8] and be able to work with them. In this work we consider the special, but the most widespread case of a multilayer waveguide.

There are the three types of waveguide modes in a regular planar optical waveguide: guided modes, substrate radiation modes, and cover radiation modes. The regular waveguide consists of a dielectric waveguide layer (or a few ones) of refractive index n_f (or n_{f1}, ..., n_{fN}) and the dielectric cladding with smaller refraction indices: n_s in the substrate layer and n_c in the cover layer. We will use Cartesian coordinates associated with the waveguide geometry. The waveguide layer thickness, say d, is about of the monochromatic electromagnetic radiation wavelength, while thicknesses of the substrate and cover layer are supposed to be much greater and, in our model, will be considered as infinite quantities.

The mathematical model of light propagation in a waveguide consists of the Maxwell equations supplemented by the matter equations and boundary conditions. In the coordinates adapted to the waveguide geometry as in Figure 1, the Maxwell equations can be split into two independent sets for the TE and TM polarizations. Their solutions are, respectively,

$$E_y(x, y, z, t) = E_y(x) \exp\{i\omega t - i\beta z\}$$

and

$$H_y(x, y, z, t) = H_y(x) \exp\{i\omega t - i\beta z\},$$

where ω is the angular frequency, β is the phase delay coefficient of the waveguide mode, x, y, z are space dimensionless coordinates, and the functions $E_y(x)$ and $H_y(x)$ for TE and TM modes, respectively, are determined by the corresponding equations

$$\frac{d^2 E_y}{dx^2} + n^2(x) E_y = \beta^2 E_y,$$

$$\frac{d^2 H_y}{dx^2} + n^2(x) H_y = \beta^2 H_y.$$

Both equations for the modes can be written in the more customary form

$$-\frac{d^2\psi}{dx^2}(k, x) + V(x)\psi(k, x) = k^2\psi(k, x). \tag{1}$$

Here $V(x) = -n^2(x)$ is a piecewise constant function (constant in each of the layer), $k^2 = -\beta^2$ is the spectral parameter, and $\psi(x) = E_y(x)$ or $H_y(x)$.

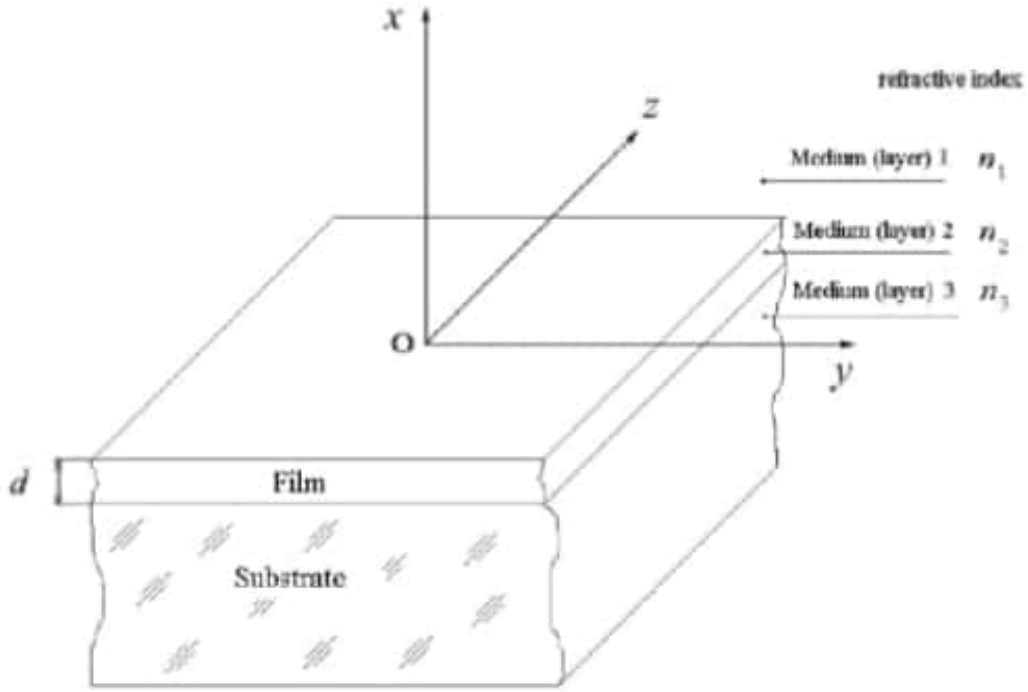


Figure 1: Waveguide is formed by media 1–3. The figure indications are: 1 is a framing medium or cover layer (air) with refractive index n_c; 2 is a waveguide layer (film) with a refractive index n_f; 3 is a substrate with refractive index n_s; d is the thickness of the waveguide layer. Film and substrate are homogeneous in the y and z directions, the substrate is usually much thicker than the film.

2. Formulation of the problem

Assumptions:

- A planar dielectric waveguide consists of homogeneous layers of isotropic materials, and the boundaries between the layer media are ideal and parallel to the xy plane.
- Electromagnetic radiation propagates in the longitudinal horizontal direction (along the z-axis) and is invariant along the transverse horizontal direction (along the y-axis).
- Electromagnetic radiation in the waveguide is monochromatic (harmonic time dependence).
- Electromagnetic radiation, for simplicity, is assumed to be linearly polarized.

The Maxwell equation in Cartesian coordinates associated with the waveguide geometry has the form (in the Gaussian units)

$$rot\vec{H} = \frac{1}{c}\frac{\partial\vec{D}}{\partial t}, \quad rot\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t}, \tag{2}$$

and can be split into the two independent sets of linear ordinary differential equations

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = ik_0\varepsilon E_y, \quad H_x = -\frac{1}{ik_0\mu}\frac{\partial E_y}{\partial z}, \quad H_z = \frac{1}{ik_0\mu}\frac{\partial E_y}{\partial x}, \tag{3}$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = ik_0\mu H_y, \quad E_x = \frac{1}{ik_0\varepsilon}\frac{\partial H_y}{\partial z}, \quad E_z = -\frac{1}{ik_0\varepsilon}\frac{\partial H_y}{\partial x}, \tag{4}$$

where $k_0 = \omega/c$ is the vacuum wave number, c is the speed of light in vacuum, ε and μ are, respectively, the dielectric constant and magnetic permeability, and $n^2 = \varepsilon\mu$ is the squared refractive index of a medium. In chosen coordinates, the boundary conditions

$$\vec{E}^{\tau}\Big|_1 = \vec{E}^{\tau}\Big|_2, \quad \vec{H}^{\tau}\Big|_1 = \vec{H}^{\tau}\Big|_2 \tag{5}$$

can be reduced to those for, respectively, TE and TM modes:

$$E_y|_1 = E_y|_2, \quad H_z|_1 = H_z|_2\,, \tag{6}$$

$$H_y|_1 = H_y|_2, \quad E_z|_1 = E_z|_2\,. \tag{7}$$

Solutions of the equations (3)–(6) and (4)–(7) yield vertical (along x axis) distributions of electromagnetic field for TE and TM modes, respectively. As functions of all the spacetime coordinates, electromagnetic fields of the modes can be written in the form

$$\begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix}(x,y,z,t) = \begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix}(x)\exp\{i\omega t - ik_0\beta z\}\,. \tag{8}$$

Transforming (3) and (4) to the form

$$\frac{d^2E_y}{dx^2} + k_0^2\left(\varepsilon\mu - \beta^2\right)E_y(x) = 0, \quad H_z = \frac{1}{ik_0\mu}\frac{dE_y}{dx}, \quad H_x = -\frac{\beta}{\mu}E_y, \tag{9}$$

$$\varepsilon\frac{d}{dx}\left(\frac{1}{\varepsilon}\frac{dH_y}{dx}\right) + k_0^2\left(\varepsilon\mu - \beta^2\right)H_y(x) = 0, \quad E_z = -\frac{1}{ik_0\varepsilon}\frac{\partial H_y}{\partial x}, \quad E_x = \frac{\beta}{\varepsilon}H_y, \tag{10}$$

both the sets can be written in the more customary form

$$-\frac{d^2\psi}{dx^2}(k,x) + V(x)\,\psi(k,x) = k^2\psi(k,x)\,. \tag{11}$$

Here $V(\tilde{x}) = -n^2(k_0x) = -\varepsilon(k_0x)\,\mu$ is a piecewise constant function (constant in each of the layer), $k^2 = -\beta^2$ is the spectral parameter, $\psi(\tilde{x}) = E_y(x)$ or $H_y(x)$, and $\tilde{x} = 2\pi(x/\lambda_0)$ is a dimensionless variable. Later on we use the notation x instead of the $\tilde{x}$.

The boundary conditions (6) and (7) hold for the function $\psi(\tilde{x})$ and its 'derivative'

$$\phi(\tilde{x}) = \frac{dE_y(x)}{dx} \quad \text{or} \quad \frac{1}{n^2(x)}\frac{dH_y(x)}{dx}\,,$$

so that

$$\psi|_1 = \psi|_2\,, \quad \phi|_1 = \phi|_2\,. \tag{12}$$

The problem of finding waveguides modes is thus reduced to the problem (11)–(12) with a potential $V(x)$ for eigenvalues k and eigenfunction $\psi(k,x)$ obeying the asymptotic conditions (Figure 2)

$$V(x) \xrightarrow[x\to-\infty]{} V_-\,, \quad V(x) \xrightarrow[x\to\infty]{} V_+\,. \tag{13}$$

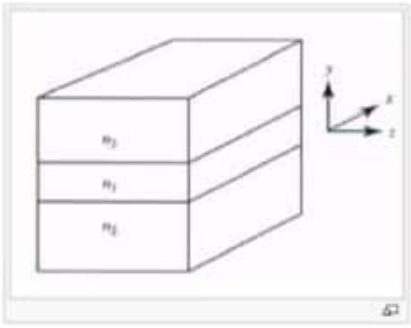
Figure 2: A schematic diagram of the potential.

In the problem (11)–(13), the operator spectrum consists of:

- a finite number of eigenvalues of the discrete spectrum $k_j = i\kappa_j$: $k_j^2 \in (\min V(x)\,,\ \min(V_-,\ V_+))$ and the corresponding eigenfunctions (guided modes);
- a continuous nondegenerate spectrum k_- : $k_-^2 \in (V_-,\ \infty)$ and the corresponding generalized eigenfunctions (substrate radiation modes);
- a continuous nondegenerate spectrum k_+ : $k_+^2 \in (V_+,\ \infty)$ and the corresponding generalized eigenfunctions (cover radiation modes).

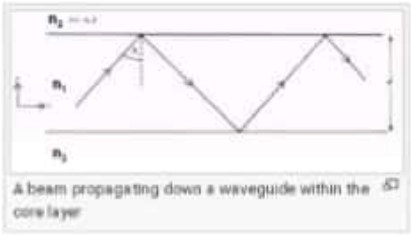
In a multilayer waveguide with a piecewise constant potential $V(x)$, solutions to the problem (11)–(13) (in the notation of (9)–(10)) in the space of square integrable functions, that is, in the case of discrete spectrum $k_j = i\kappa_j$, were considered in a large number of research studies, both theoretical and computational. There are basic studies [9, 10, 11] and reviews [12, 13, 14, 15, 16] on integrated optics devoted to guided modes in waveguides. The pioneering [9, 10, 11] and recent works [17, 18, 19] on integrated optics are devoted to numerical methods of constructing the function $\psi_j(x)$ as a linear combination of a fundamental system of solutions of the equation (11) in each of the layers, with subsequent matching these functions at the layer interfaces according to (12).

Planar dielectric waveguides



Planar (slab) waveguides are the basis of waveguides used in integrated optoelectronics. The same mathematical ideas can be applied (with minor modifications) to circular waveguides.

The waveguide consists of a semi-infinite slab of dielectric materials with thickness d and refractive index n_1 (the core) that is sandwiched between two regions (the cladding) both of refractive index n_2, and where $n_1 > n_2$.



A beam propagating down a waveguide within the core layer

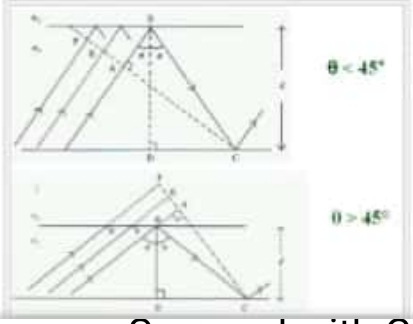
A ray of light may propagate down the core provided that total internal reflection occurs at the core/cladding interface.

this requires that:

$$90\,\mathrm{deg} > \theta_i > \theta_c$$

Where

θ_i is the internal ray angle (from now on written as θ)



In fact there are 'infinite" number of rays, all slightly displaced from each other, also propagating down the

The total phase change is equivalent to:

$$(AB + BC)\frac{2\pi n_1}{\lambda_0} - 2\phi(\theta)$$

Where

λ_0 is the wavelength of light in the medium

To determine the path of the light from a to b to c using trigonometry:

$$AB = BC\cos 2\theta$$

Thus

$$AB + BC = BC(1 + \cos 2\theta)$$

Since

$$\cos 2\theta = 2\cos^2\theta - 1$$
$$AB + BC = 2BC\cos^2\theta$$

$BD = BC\cos\theta = d$ that is the thickness of the slab

So that

$$AB + BC = 2d\cos\theta$$

The thickness of the slab determines the number of modes or angles that light will propagate at.

In order for the mode to propagate the total phase change must be a multiple of 2π:

$$\frac{4\pi n_1 d\cos\theta}{\lambda_0} - 2\phi(\theta) = 2m\pi$$
$$\frac{2\pi n_1 d\cos\theta}{\lambda_0} - \phi(\theta) = m\pi$$

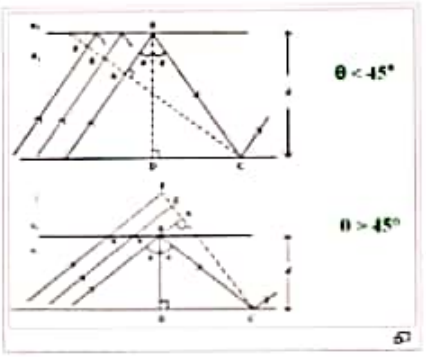
Where

m is an integer

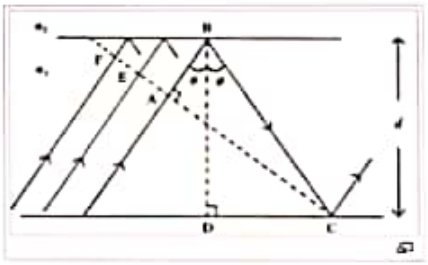
So for each value of m there will be an angle θ_m that satisfies the equation

Each value of θ_m (those $>\theta_c$) has a distinct distribution of electric field across the guide. This distribution is known as a mode. Depending

In fact there are 'infinite" number of rays, all slightly displaced from each other, also propagating down the guide. The dotted line that is perpendicular to the wave lines is the wavefront of the propagating beam. The rays represent lines drawn normally to the plane wavefronts.



The wavefront FC intersects tow the upwardly traveling portions of the same ray at points A and C. Therefore the phase at C and A must be the same or differ by a multiple of 2π.



Otherwise there would be destructive interference between out-of-phase waves and the light will not propagate. It also requires very specific angles θ above the critical angle.

Consider the phase difference between A and C

There are two factors -the path length of AB + BC -the phase change due to reflection at B and C

We write the phase change resulting from reflection simply as δ(θ)

For perpendicular radiation φ (θ) is 2 ψ, for parallel radiation φ(θ) = 2 δ.

11-6: Dielectric Waveguides

In the case of a metallic waveguide we assumed that there was no field in the conductors. Therefore we needed to solve Maxwell's Equations only for the region inside the cavity. Let us now consider the situation where we have an electromagnetic wave confined to a dielectric system whose medium has index of refraction n_1 and which is surrounded by a region of index of refraction n_2. We require that $n_1 > n_2$ so that the process of confinement is due to total internal reflection not reflection off a conducting surface. Thus the reflection theoretically is totally lossless whereas this is not so for a metal waveguide at ordinary temperatures. However, the dielectric waveguide will lose energy due to absorption in the medium.

Dielectric waveguides have increasingly important applications for modern communications because of the emerging technology of fiber optics. Thus the study of dielectric waveguides is of current interest because it is becoming the transmitter of choice for infrared and visible frequencies. We have noted that coaxial cables are effective up to 100 MHz. Above that resistive losses make it ineffective. We then must go to waveguides which are effective into the 10 GHz region. Again resistive losses begin to take over. Our next option, therefore, is to use the dielectric waveguides which are effective into the 10 THz region. Of more importance, however, is the bandwidth capability of these devices. If we assume that a 5% bandwidth is reasonable for each device we note that for coaxial cables this is 5 MHz, for waveguides this is 500 MHz, but for optical fibers this is 500 GHz. Since the amount of data which can be transmitted is related to the bandwidth we can see that there is significant potential for optical fibers in communications is enormous.

Example 11-5: The Dielectric Slab

In a dielectric waveguide we need the solution of Maxwell's Equations in both the internal and external dielectrics and use the appropriate continuity relations at

Each value of θ_m (those $>\theta_c$) has a distinct distribution of electric field across the guide. This distribution is known as a mode. Depending on the mode there may a distribution that is centered in the core or may have 2 spots, 4 spots etc when view in cross section.

When: $\theta_m is = \theta_c$: the mode is at cut-off

If

$\theta_m < \theta_c$: the mode is below cutoff resulting in rapid attenuation and light will not be propagated.

If

$\theta_m > \theta_c$: the mode is above cut-off which can propagate

Mode Cut-off

The mode will be at cut-off if

$$\frac{2}{\pi} n_1 cos\theta_c \lambda_0 = \pi m$$

Since

$$cos\theta_c = (1 - sin2\theta_c)^{1/2} = [1 - (n_2/n_1)\,2]^{1/2}$$
$$sin\theta_c = (n_2/n_1)$$

Therefore mode cut-off occurs when:

$$\frac{2\pi d n_1}{\lambda_0}\left[1 - (\frac{n_2}{n_1})^2\right]^{1/2} = \pi m$$

Or

$$V = \frac{\pi}{2} m$$

Where

V is the normalized film thickness, normalized frequency, or V parameter

D is the thickness of the waveguide

Where

V is the normalized film thickness, normalized frequency, or V parameter

D is the thickness of the waveguide

Thus the mode shift depends on the relative index of refraction of the materials and the thickness of the waveguide. If you decrease the thickness you decrease the number of modes that can propagate.

The number of guided modes in a step-index guide (N) is given by

$$N = 1 + int(2V/\pi)$$

Only one mode propagates if $V < \frac{\pi}{2}$, in this case it is called a single mode guide.

$$\frac{d}{\lambda_0} < \frac{1}{2(n_1^2 n_2^2)^{1/2}}$$

Example

Find the maximum waveguide core thickness d for a device with n_1 =1.48 and n_2 = 1.46 and a wavelength of 1 μ m to achieve single mode propagation.

$$d < \frac{1}{2(1.48^2 1.46^2)^{1/2}} \mu m$$

$$V < 2.06 \mu m$$

$$\mathcal{E}_z = A_1 \sin \gamma_1 y + A_2 \cos \gamma_1 y \tag{11-409}$$

We choose the conditions such that:

$$\mathcal{E}_z = A_2 \cos \gamma_1 y \tag{11-410}$$

is the description of the wave in the region, $-b/2 \leq y \leq b/2$. This solution makes $\mathcal{E}_z$ an even function with respect to the midplane of the slab.

Because we are under conditions of total internal reflection outside the dielectric interface $\gamma < 0$ and we let $\gamma^2 = -\beta^2$. Our solution is now:

$$\mathcal{E}_z = A_3 e^{-\beta y} \qquad y \geq b/2 \tag{11-411}$$

$$\mathcal{E}_z = A_4 e^{\beta y} \qquad y \leq -b/2 \tag{11-412}$$

where:

$$\beta^2 = \kappa_g^2 - \frac{\omega^2}{c^2} \tag{11-413}$$

Equations 11-408 and 11-413 place conditions on the range of values of κ_g. For example, for wave motion to exist inside the slab, we must have $\gamma_1 > 0$. Thus:

$$n_0^2 \frac{\omega^2}{c^2} - \kappa_g^2 \geq 0 \tag{11-414}$$

and

$$\kappa_g < n_0 \frac{\omega}{c} \tag{11-415}$$

Also for there to be a decaying field outside the slab, we must have $\beta > 0$. Thus:

$$\kappa_g^2 - \frac{\omega^2}{c^2} > 0 \tag{11-416}$$

and the value of κ_g lies in the range:

$$\frac{\omega}{c} < \kappa_g < n_0 \frac{\omega}{c} \tag{11-417}$$

Using Equations 11-39 through 11-43, we derive the fields inside the dielectric slab:

$$\mathcal{E}_x = \frac{i}{\gamma_1^2}\left(\kappa_g \frac{\partial \mathcal{E}_z}{\partial x}\right) = 0 \tag{11-418}$$

$$\mathcal{E}_y = \frac{i}{\gamma_1^2}\left(\kappa_g \frac{\partial \mathcal{E}_z}{\partial y}\right) = -\frac{i}{\gamma_1^2} \kappa_g A_2 \gamma_1 \sin \gamma_1 y = -\frac{i\kappa_g}{\gamma_1} A_2 \sin \gamma_1 y \tag{11-419}$$

$$\mathcal{E}_z = A_2 \cos \gamma_1 y \tag{11-420}$$

$$\mathcal{B}_x = -\frac{i}{\gamma_1^2} \frac{n_0^2 \omega}{c^2} \frac{\partial \mathcal{E}_z}{\partial y} = \frac{i}{\gamma_1^2} \frac{n_0^2 \omega}{c^2} \gamma_1 A_2 \sin \gamma_1 y = \frac{i}{\gamma_1} \frac{n_0^2 \omega}{c^2} A_2 \sin \gamma_1 y \tag{11-421}$$

$$\mathcal{B}_y = \frac{i}{\gamma_1^2} \frac{n_0^2 \omega}{c^2} \frac{\partial \mathcal{E}_z}{\partial x} = 0 \tag{11-422}$$

$$\mathcal{B}_z = 0 \tag{11-423}$$

This above combination, the TM mode, and $\mathcal{E}_z = A_2 \cos \gamma_1 y$, results in fields which are known as the even TM modes. The combination, the TM mode, and $\mathcal{E}_z = A \sin \gamma_1 y$, results in fields which are known as the odd TM modes. Odd and even combinations of the TE modes are also possible.

We now derive the conditions for wave transmission in the slab. We know that the tangential component of $\mathcal{E}$, which in our case is $\mathcal{E}_z$, is continuous across the boundary. Thus, at $y = b/2$:

Problem 11-28

Find the expression for the fields for the dielectric slab under the conditions that the medium is air (n = 1.0) on one side and a conductor on the other. The index of refraction of the slab is n. Find an expression for the possible modes for this waveguide and the cut-off frequency. Compare with the four possible modes of the symmetric dielectric slab waveguide.

Example 11-6: The Circular Optical Fiber

We next discuss the more complicated example of an optical fiber. We will assume that it is a circular cylinder and that it consists of a core of radius a and a medium of dielectric constant n_1 surrounded by a cylindrical sleeve called cladding, which has dielectric constant n_2, as we have illustrated in Figure 11-15. We will

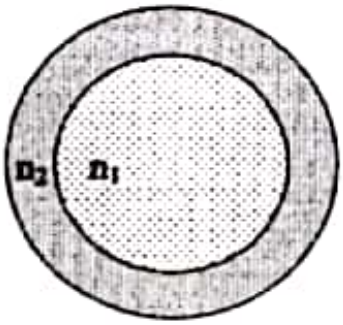


Figure 11-15: The Optical Fiber

consider only situations where n_1 is greater than n_2. We find the forms of the waves inside the fiber using the equations noted in Section 11-2, but we take into consideration the fact that the waves are traveling in a dielectric. Thus:

$$\mathcal{E}_\rho = \frac{i}{\gamma'^2}\left[\kappa_g \frac{\partial \mathcal{E}_z}{\partial \rho} + \frac{\omega}{\rho}\frac{\partial \mathcal{B}_z}{\partial \phi}\right] \tag{11-447}$$

$$\mathcal{E}_\phi = \frac{i}{\gamma'^2}\left[\frac{\kappa_g}{\rho}\frac{\partial \mathcal{E}_z}{\partial \phi} - \omega\frac{\partial \mathcal{B}_z}{\partial \rho}\right] \tag{11-448}$$

$$\mathcal{B}_\rho = \frac{i}{\gamma'^2}\left[\kappa_g \frac{\partial \mathcal{B}_z}{\partial \rho} - n_0^2\frac{\omega}{c^2}\frac{1}{\rho}\frac{\partial \mathcal{E}_z}{\partial \phi}\right] \tag{11-449}$$

$$\mathcal{B}_\phi = \frac{i}{\gamma'^2}\left[\frac{\kappa_g}{\rho}\frac{\partial \mathcal{B}_z}{\partial \phi} + n_0^2\frac{\omega}{c^2}\frac{1}{\rho}\frac{\partial \mathcal{E}_z}{\partial \rho}\right] \tag{11-450}$$

Here n_0 is the index of refraction of the material and γ' is given by:

$$\gamma'^2 = n_0^2\frac{\omega^2}{c^2} - \kappa_g^2 \tag{11-451}$$

We have shown that $\mathcal{E}_z$ and $\mathcal{B}_z$ will be the solution to Bessel's Equation. For the inside cylinder the general solution is:

$$\mathcal{E}_z = [A_1 J_m(\gamma_1\rho) + B_1 N_m(\gamma_1\rho)]\left[C_1 e^{im\phi} + D_1 e^{-im\phi}\right] e^{i(\kappa_g z - \omega t)} \tag{11-452}$$

Taking the limiting value of γ_1 Equation 11-436 becomes:

$$\sqrt{(n_0^2 - 1)\frac{\omega_c^2}{c^2}}\ \frac{b}{2} = \left(m - \frac{1}{2}\right)\pi \tag{11-437}$$

which allows us to find the possible values for ω_c. Rearranging Equation 11-437 we have:

$$\omega_c = \frac{(m-1/2)}{\sqrt{(n_0^2-1)}}\ \frac{2\pi c}{b} \tag{11-438}$$

Thus:

$$\nu_c = \frac{(m-1/2)}{\sqrt{n_0^2-1}}\ \frac{c}{b} \tag{11-439}$$

As an example suppose we have a glass slab of index of refraction 1.5 and thickness 0.4 cm . The first cut-off frequency is given by:

$$\nu_c = \frac{1}{2}\ \frac{1}{\sqrt{2.25-1}}\ \frac{3\times10^8\ \text{m-s}^{-1}}{4\times10^{-3}\ \text{m}} =$$

$$\frac{3\times10^{11}}{1.12\times4}\ \text{Hz} = 6.7\times10^{10}\ \text{Hz} = 67\ \text{GHz} \tag{11-440}$$

We next wish to find out the parameters of the wave at a frequency above threshold. We are given the quantities: the thickness of the slab, b; the index of refraction of the slab, n_0; and the frequency of the signal we want to transmit, ω. What we do not know usually is the propagation constant, κ_g, and thus the wavelength of the signal in the guide. We find this from Equation 11-431.

We have:

$$\left(\gamma_1\ \frac{b}{2}\right)^2 + \left(\beta\ \frac{b}{2}\right)^2 = (n_0^2 - 1)\left(\frac{\omega b}{2c}\right)^2 \tag{11-441}$$

where we have multiplied through by b/2. Next we multiply Equation 11-432 through by the common factor of b/2, getting:

$$\left(\frac{\gamma_1 b}{2}\right)\cot\left(\frac{\gamma_1 b}{2}\right) = -n_0^2\left(\frac{\beta b}{2}\right) =$$

$$-n_0^2\sqrt{(n_0^2-1)\left(\frac{\omega b}{2c}\right)^2 - \left(\frac{\gamma_1 b}{2}\right)^2} \tag{11-442}$$

where the last term comes from Equation 11-441.

We define a variable X as:

$$X = \frac{\gamma_1 b}{2} \tag{11-443}$$

and Equation 11-442, when rearranged, becomes:

$$\cot(X) = -\frac{n_0^2}{X}\sqrt{(n_0^2-1)\left(\frac{\omega b}{2c}\right)^2 - X^2} =$$

$$-n_0^2\sqrt{\frac{(n_0^2-1)(\omega b/2c)}{X^2} - 1} \tag{11-444}$$

$$= -n_0^2\sqrt{\frac{X_0^2}{X^2} - 1} = F(X) \tag{11-445}$$

This equation does not have a solution determined by ordinary functions. However, it can be solved by graphing the two functions. Thus we now plot F(X) and cot (X) on the same set of coordinates. The cotangent function consists of a series of curves ranging from a value of $\cot(X) = \infty$ at $X = 0$; $\cot(X) = 0$ at $\pi/2$;

satisfied at point E in Fig. 3. (The corresponding point at $\phi = 45°$ in Fig. 1 is not marked.) Figure 3 also clearly shows that $\Delta = 90°$ is now possible at the two points marked D_1 and D_2. The corresponding angles of incidence ϕ_1 and ϕ_2 are determined by[10]

$$\sin^2 \phi_{1,2} = (\tfrac{1}{4})[(n^2 + 1) \mp (n^4 - 6n^2 + 1)^{1/2}], \quad (16)$$

where $n = 1/N$, as before. For the Ge–air interface $n = 0.25$, and Eq. (16) gives $\phi_1 = 15.04°$ (which is slightly above the critical angle $\phi_c = 14.48°$) and $\phi_2 = 42.93°$. A Ge, single-reflection, quarter-wave retarder can be designed to operate at $\phi_2 = 42.93°$ (point D_2 in Fig. 3) with acceptable angular sensitivity. Standard silica-based Fresnel rhomb quarter-wave retarders use 2, 3, or 4 TIRs.[11]

At the angle of incidence $\phi_m = \phi_a$, where $\Delta = \Delta_{max}$, we can solve the following two equations:

$$(\delta_p/2) + (\delta_s/2) = \pi/2,$$
$$(\delta_p/2) - (\delta_s/2) = \Delta_{max}/2, \quad (17)$$

to obtain δ_p and δ_s in terms of Δ_{max} only:

$$\delta_p = (\pi + \Delta_{max})/2,$$
$$\delta_s = (\pi - \Delta_{max})/2. \quad (18)$$

Consequently, the reflection Jones matrix[2] at $\phi_m = \phi_a$ takes the simple form

$$R = j\begin{bmatrix} \exp(j\Delta_{max}/2) & 0 \\ 0 & \exp(-j\Delta_{max}/2) \end{bmatrix}. \quad (19)$$

The outside multiplier j in Eq. (19) represents the quarter-wave average phase shift on reflection.

3. EQUAL AVERAGE AND DIFFERENTIAL PHASE SHIFTS

At the point F where the δ_s and Δ-versus-ϕ curves intersect in Fig. 3, the average and differential reflection phase shifts are equal,

$$\delta_a = \Delta,$$
$$(\delta_p + \delta_s)/2 = \delta_p - \delta_s,$$
$$\delta_p = 3\delta_s. \quad (20)$$

Equations (20), which are to our knowledge new, bear some resemblance to the Abelès condition,[3] $\Delta = \delta_s$, or $\delta_p = 2\delta_s$, which occurs at $\phi = 45°$. However, the angle of incidence at which Eq. (20) is satisfied is a function of N given by

$$\sin^2 \phi = (N^2 + 1)/(4N^2). \quad (21)$$

The derivation leading to Eq. (21), starting from Eqs. (4) and (7), is a bit lengthy and requires some patience, so it is omitted here to save space. Equation (21) has an acceptable solution for ϕ only if

$$\sin^2 \phi \geq \sin^2 \phi_c = 1/N^2. \quad (22)$$

Based on Eqs. (21) and (22), an acceptable nontrivial solution[12] of Eq. (20) exists only if

$$N > \sqrt{3} = 1.732. \quad (23)$$

For the Ge–air interface $N = 4$, and Eq. (21) gives $\phi = 31.02°$, which precisely locates the point F in Fig. 3.

4. LIMITING SLOPES OF THE Δ-versus-ϕ CURVE AT THE CRITICAL ANGLE AND GRAZING INCIDENCE

By taking the derivative of both sides of Eq. (4) with respect to ϕ, we obtain

$$\tfrac{1}{2}[\sec^2(\Delta/2)](\partial\Delta/\partial\phi) = [(1 - N^2)\tan^2 \phi + 2]/[N \sin \phi \tan^2 \phi \times (N^2 \sin^2 \phi - 1)^{1/2}]. \quad (24)$$

The maximum differential reflection phase shift Δ occurs at the angle ϕ_m where $\partial\Delta/\partial\phi = 0$. By setting the numerator of the right-hand side of Eq. (24) equal to 0, we get

$$\tan^2 \phi_m = 2/(N^2 - 1), \quad (25)$$
$$\sin^2 \phi_m = 2/(N^2 + 1). \quad (26)$$

Equations (11) and (26) lead to the conclusion that $\phi_m = \phi_a$, as has already been noted in Section 2.[13]

At the critical angle, $N^2 \sin \phi = 1$, and the slope of the Δ-versus-ϕ curve is infinite,

$$(\partial\Delta/\partial\phi)_{\phi=\phi_c} = \infty, \quad (27)$$

as can be seen from Eq. (24). Similarly, by differentiation of Eqs. (2) and (3), it can also be verified that

$$(\partial\delta_p/\partial\phi)_{\phi=\phi_c} = (\partial\delta_s/\partial\phi)_{\phi=\phi_c} = \infty. \quad (28)$$

Therefore all phase shifts rise vertically with respect to ϕ at the critical angle. (The angular resolution used in plotting Figs. 1 and 3 does not make this clear.)

The limiting slope of the Δ-versus-ϕ curve at grazing incidence ($\phi = 90°$) is also obtained from Eq. (24) as

$$(\partial\Delta/\partial\phi)_{\phi=90°} = -(2/N)(N^2 - 1)^{1/2} = -2(1 - n^2)^{1/2} = -2\cos\phi_c. \quad (29)$$

For large N, the limiting slope of the Δ-versus-ϕ curve at grazing incidence in TIR approaches -2. For the Ge–air interface ($N = 4$), the limiting slope equals -1.9365.

We have also verified analytically that the second derivative of Δ with respect to ϕ at grazing incidence is zero,

$$(\partial^2\Delta/\partial\phi^2)_{\phi=90°} = 0. \quad (30)$$

Therefore, Δ is nearly a linear function of ϕ near grazing incidence, as can be readily inferred from Figs. 1 and 3. In terms of the grazing incidence angle

$$\theta = 90° - \phi, \quad (31)$$

Δ can be expressed as

$$\Delta \approx 2(1 - n^2)\,\theta. \quad (32)$$

The approximation in Eq. (32) is good for small θ, and gets better as N increases. For example, when $n = 0.25$ ($N = 4$) and $\theta = 10°$, Eq. (32) gives $\Delta = 19.365°$, which is only $0.008°$ above the exact value predicted by Eq. (4).

Equation (32) has potential application for describing changes in the state of polarization of light that travels as

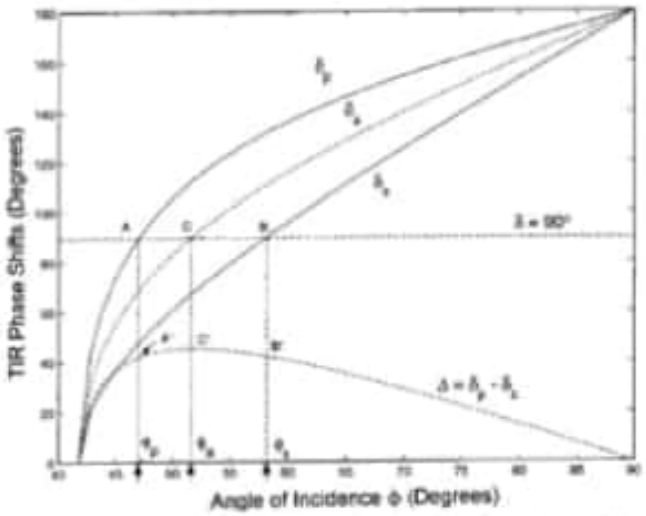


Fig. 1. TIR phase shifts δ_p, δ_s, δ_a, and Δ at the glass–air interface ($N = 1.5$) plotted as functions of the angle of incidence ϕ between the critical angle $\phi_c = \arcsin(2/3) = 41.81°$ and grazing incidence $\phi = 90°$. Quarter-wave phase shifts are attained at points A, B, and C where the line $\delta = 90°$ intersects the curves of δ_p, δ_s, and δ_a at angles of incidence denoted by ϕ_p, ϕ_s, and ϕ_a, respectively.

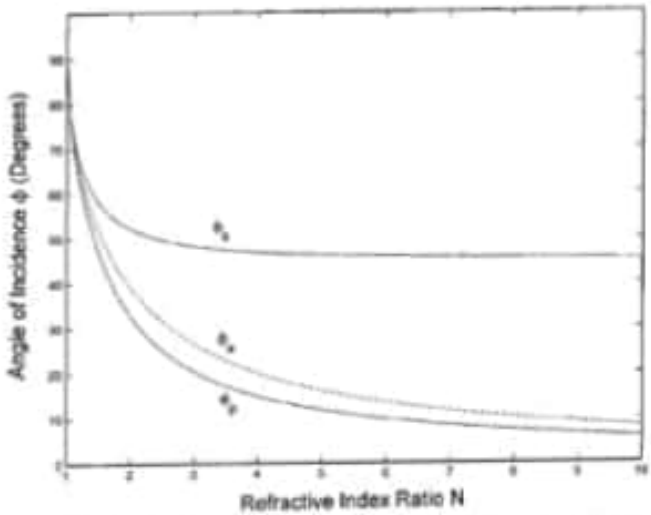


Fig. 2. Three angles of incidence ϕ_p, ϕ_s, and ϕ_a, at which the p, s, and average TIR phase shifts [Eqs. (9)–(11)] are quarter-wave are plotted as functions of the refractive index ratio N.

$= 1.5$) plotted as functions of the angle of incidence ϕ between the critical angle $\phi_c = \arcsin(2/3) = 41.81°$ and grazing incidence $\phi = 90°$. Quarter-wave phase shifts are attained at points A, B, and C where the line $\delta = 90°$ intersects the curves of δ_p, δ_s, and δ_a at angles of incidence denoted by ϕ_p, ϕ_s, and ϕ_a, respectively.

Expressions for the angles ϕ_p and ϕ_s for any N are readily obtained by setting the left-hand sides of Eqs. (2) and (3) equal to 1. The corresponding expression for ϕ_a is obtained by setting the denominator of the right-hand side of Eq. (7) equal to 0. The results are listed below:

$$\sin^2\phi_p = (N^2 + 1)/(N^4 + 1), \tag{9}$$

$$\sin^2\phi_s = (N^2 + 1)/(2N^2), \tag{10}$$

$$\sin^2\phi_a = 2/(N^2 + 1). \tag{11}$$

Figure 2 shows the three angles ϕ_p, ϕ_s, and ϕ_a plotted as functions of N. All angles approach 90° as N tends to 1 and decrease monotonically as N is increased. For large N, ϕ_s approaches 45°, whereas ϕ_p and ϕ_a approach 0.

It will be shown in Section 4 that the angle ϕ_a for quarter-wave average phase shift given by Eq. (11) is exactly the same as the angle $\phi_{\Delta m}$ at which Δ is maximum [see Eq. (26) below]. (In Fig. 1 the peak point C′ of the Δ-versus-ϕ curve lies vertically below point C.) Curiously, this also happens to be the angle of incidence at which p- and s-polarized light tunnel equally across a thin uniform air gap between two half-prisms of the same refractive index under conditions of frustrated TIR.[6–8]

It is also of interest to calculate δ_s when $\delta_p = \pi/2$, and δ_p when $\delta_s = \pi/2$. The results are readily obtained from Eq. (8):

$$\tan(\delta_s/2) = 1/N^2, \qquad \delta_p = \pi/2,$$

$$\tan(\delta_p/2) = N^2, \qquad \delta_s = \pi/2. \tag{12}$$

Consequently, the differential reflection phase shifts at ϕ_p and ϕ_s are

$$\Delta(\phi_p) = (\pi/2) - 2\arctan(1/N^2),$$

$$\Delta(\phi_s) = 2\arctan(N^2) - (\pi/2). \tag{13}$$

From Eqs. (13) it follows that

$$\Delta(\phi_p) = \Delta(\phi_s). \tag{14}$$

The equal differential reflection phase shifts are represented by points A′ and B′ in Fig. 1.

The maximum differential reflection phase shift (at point C′) is given by[9]

$$\tan(\Delta_{max}/2) = (N^2 - 1)/(2N). \tag{15}$$

When $N = 1.5$, $\Delta_{max} = 45.24°$ and the Δ-versus-ϕ curve lies entirely below the $\delta = 90°$ line in Fig. 1. To achieve quarter-wave differential phase shift on single reflection (i.e., $\Delta = 90°$), we must have $N \geq \sqrt{2} + 1 = 2.414$, according to Eq. (15).

Figure 3 is similar to Fig. 1 except that now $N = 4$, which corresponds to the Ge–air interface in the infrared. The Abelès condition,[5] $\Delta = \delta_s$, or $\delta_p = 2\delta_s$ at $\phi = 45°$, is

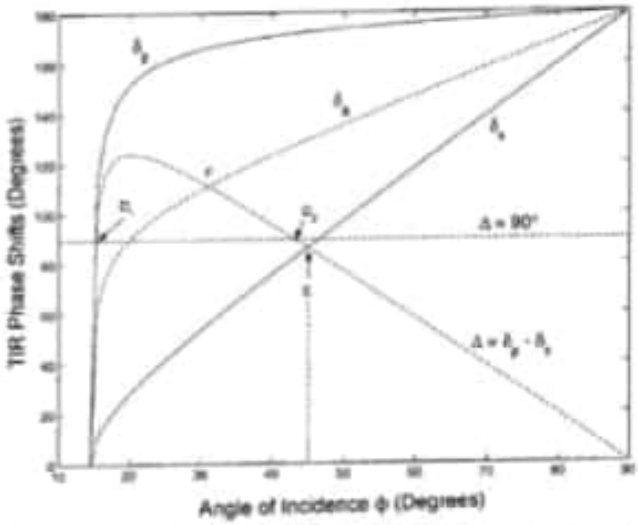


Fig. 3. As in Fig. 1 except that $N = 4$, which corresponds to the Ge–air interface in the infrared. The significance of the marked points D_1, D_2, E, and F is discussed in the text.

Phase shifts that accompany total internal reflection at a dielectric–dielectric interface

R. M. A. Azzam

Department of Electrical Engineering, University of New Orleans, New Orleans, Louisiana 70148

Received January 13, 2004; revised manuscript received March 18, 2004, accepted March 23, 2004

The absolute, average, and differential phase shifts that p- and s-polarized light experience in total internal reflection (TIR) at the planar interface between two transparent media are considered as functions of the angle of incidence ϕ. Special angles at which quarter-wave phase shifts are achieved are determined as functions of the relative refractive index N. When the average phase shift equals $\pi/2$, the differential reflection phase shift Δ is maximum, and the reflection Jones matrix assumes a simple form. For $N > \sqrt{3}$, the average and differential phase shifts are equal (hence $\delta_p = 3\delta_s$) at a certain angle ϕ that is determined as a function of N. All phase shifts rise with infinite slope at the critical angle. The limiting slope of the Δ-versus-ϕ curve at grazing incidence $(\partial\Delta/\partial\phi)_{\phi=90^\circ} = -(2/N)(N^2 - 1)^{1/2} = -2\cos\phi_c$, where ϕ_c is the critical angle and $(\partial^2\Delta/\partial\phi^2)_{\phi=90^\circ} = 0$. Therefore Δ is proportional to the grazing incidence angle $\theta = 90^\circ - \phi$ (for small θ) with a slope that depends on N. The largest separation between the angle of maximum Δ and the critical angle is 9.88° and occurs when $N = 1.55377$. Finally, several techniques are presented for determining the relative refractive index N by using TIR ellipsometry. © 2004 Optical Society of America

OCIS codes: 240.0240, 260.0260, 260.2130, 260.5430, 260.6970.

1. INTRODUCTION

The reflection of p- and s-polarized light at the planar interface between two semi-infinite, homogeneous, isotropic media is governed by the well-known Fresnel coefficients.[1–3] When both media are transparent, the phase shifts that accompany partial external or internal reflection assume the values of 0 or π. However, under conditions of total internal reflection (TIR), at angles of incidence ϕ above the critical angle

$$\phi_c = \arcsin n, \tag{1}$$

the phase shifts δ_p, δ_s, and $\Delta = \delta_p - \delta_s$ are nontrivial and are given by[2,3]

$$\tan(\delta_p/2) = N(N^2\sin^2\phi - 1)^{1/2}/\cos\phi, \tag{2}$$

$$\tan(\delta_s/2) = (N^2\sin^2\phi - 1)^{1/2}/(N\cos\phi), \tag{3}$$

$$\tan(\Delta/2) = (N^2\sin^2\phi - 1)^{1/2}/(N\sin\phi\tan\phi). \tag{4}$$

In Eqs. (1)–(4),

$$N = N_0/N_1 \tag{5}$$

is the high-to-low ratio of refractive indices of the media of incidence and (evanescent) refraction, respectively ($N > 1$), and $n = 1/N < 1$.

In addition to the differential reflection phase shift Δ (which is measurable by ellipsometry[3]), we introduce the average phase shift on reflection

$$\delta_a = (\delta_p + \delta_s)/2. \tag{6}$$

Based on Eqs. (2) and (3) and the trigonometric identity for the tangent of the sum of two angles, we obtain

$$\tan\delta_a = [(\cos\phi)(N + N^{-1})(N^2\sin^2\phi - 1)^{1/2}]/[2 - (\sin^2\phi)(N^2 + 1)]. \tag{7}$$

From Eqs. (2) and (3), we also obtain the following useful relation between δ_p and δ_s,

$$\tan(\delta_p/2) = N^2\tan(\delta_s/2), \tag{8}$$

which is valid for any interface (a given N) at any angle of incidence in the range $\phi_c \leqslant \phi \leqslant 90^\circ$. Another direct relation between δ_p and δ_s that is valid at a given ϕ independent of N was previously published.[4]

In this paper we derive a number of interesting new results concerning the TIR phase shifts δ_p, δ_s, δ_a, and Δ. In Section 2 we find the angles of incidence at which these phase shifts are quarter-wave ($=\pi/2$) and discuss their significance. Analogous to the Abelès condition,[5] $\delta_p = 2\delta_s$ at $\phi = 45^\circ$, we investigate the condition $\delta_a = \Delta$, which is equivalent to $\delta_p = 3\delta_s$, in Section 3. In Section 4 the limiting slopes of the Δ-versus-ϕ curve at the critical angle and at grazing incidence are obtained. In Section 5, the difference between the angle of incidence ϕ_m of maximum Δ and the critical angle ϕ_c is obtained as a function of N, and it is shown that this difference $\phi_m - \phi_c$ cannot exceed 9.88°. In Section 6 we propose several methods for recovering N based on various features of the Δ-versus-ϕ curve. Section 7 is a brief summary of the paper.

2. QUARTER-WAVE PHASE SHIFTS IN TOTAL INTERNAL REFLECTION

As a specific example, Fig. 1 shows the phase shifts δ_p, δ_s, δ_a, and Δ on TIR at the glass–air interface (N

1084-7529/2004/081559-05$15.00

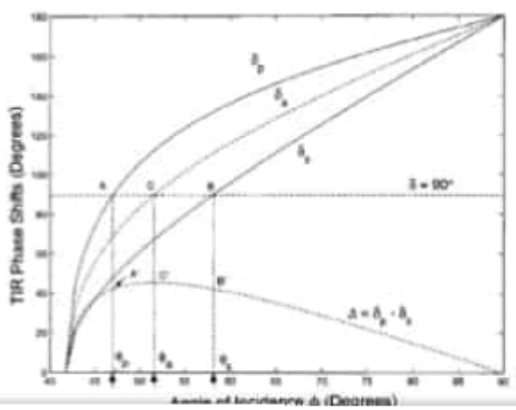


Figure 2 shows the three angles ϕ_p, ϕ_s, and ϕ_a plotted as functions of N. All angles approach 90° as N tends to 1 and decrease monotonically as N is increased. For large N, ϕ_s approaches 45°, whereas ϕ_p and ϕ_a approach 0.

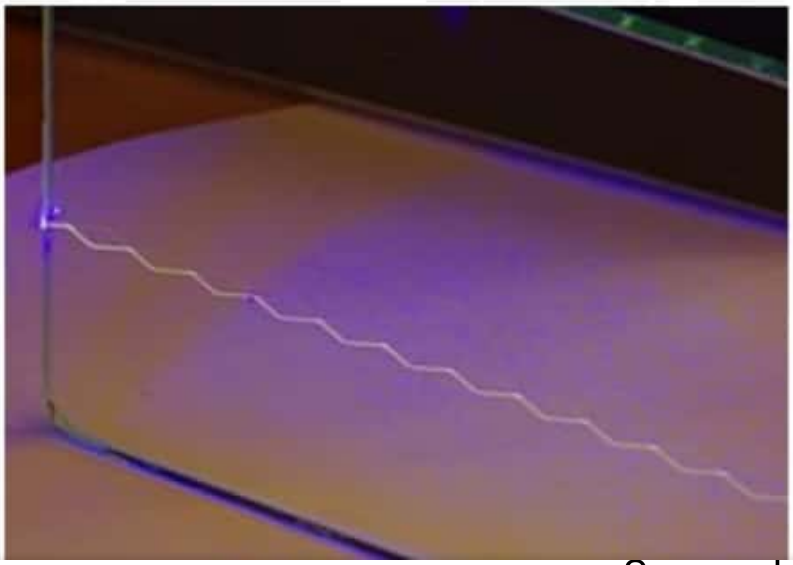
It will be shown in Section 4 that the angle ϕ_a for quarter-wave average phase shift given by Eq. (11) is exactly the same as the angle ϕ_m at which Δ is maximum [see Eq. (26) below]. (In Fig. 1 the peak point C′ of the Δ-versus-ϕ curve lies vertically below point C.) Curiously, this also happens to be the angle of incidence at which p- and s-polarized light tunnel equally across a thin uniform air gap between two half-prisms of the same refractive index under conditions of frustrated TIR.[6–8]

It is also of interest to calculate δ_s when $\delta_p = \pi/2$, and δ_p when $\delta_s = \pi/2$. The results are readily obtained from Eq. (6):

Refraction is generally accompanied by *partial* reflection. When waves are refracted from a medium of lower propagation speed to a medium of higher propagation speed (e.g., from water to air), the *angle of refraction* (between the refracted ray and the line perpendicular to the refracting surface) is greater than the *angle of incidence* (between the incident ray and the perpendicular). As the angle of incidence approaches a certain limit, called the *critical angle*, the angle of refraction approaches 90°, at which the refracted ray becomes parallel to the surface. As the angle of incidence increases beyond the critical angle, the conditions of refraction can no longer be satisfied; so there is no refracted ray, and the partial reflection becomes total. For visible light, the critical angle is about 49° for incidence at the water-to-air boundary, and about 42° for incidence at the common glass-to-air boundary.

Details of the mechanism of TIR give rise to more subtle phenomena. While total reflection, by definition, involves no continuing flow of power *across* the interface between the two media, the

Total internal reflection (**TIR**) is the optical phenomenon in which the surface of the water in a fish-tank, viewed from below the water level, reflects the underwater scene like a mirror, with no loss of brightness (Fig. 1). In general, TIR occurs when waves in one medium reach the boundary with another medium at a sufficiently slanting angle, provided that the second ("external") medium is transparent to the waves and allows them to travel faster than in the first ("internal") medium. TIR occurs not only with electromagnetic waves such as light and microwaves, but also with other types of waves, including sound and water waves. In the case of a narrow train of waves, such as a laser beam (Fig. 2), we tend to describe the reflection in terms of "rays" rather than waves. In a medium whose properties are independent of direction, such as air, water, or glass, each "ray" is perpendicular to the associated wavefronts.[*importance?*]



distance from the interface. The "total" reflection is indeed total if the external medium is lossless (perfectly transparent), continuous, and of infinite extent, but can be conspicuously *less* than total if the evanescent wave is absorbed by a lossy external medium ("attenuated total reflectance"), or diverted by the outer boundary of the external medium or by objects embedded in that medium ("frustrated" TIR). Unlike *partial* reflection between transparent media, total internal reflection is accompanied by a non-trivial phase shift (not just zero or 180°) for each component of polarization (perpendicular or parallel to the plane of incidence), and the shifts vary with the angle of incidence. The explanation of this effect by Augustin-Jean Fresnel, in 1823, added to the evidence in favor of the wave theory of light.

The phase shifts are utilized by Fresnel's invention, the Fresnel rhomb, to modify polarization. The efficiency of the reflection is exploited by optical fibers (used in telecommunications cables and in image-forming fiberscopes), and by reflective prisms, such as erecting prisms for binoculars.