Progress In Electromagnetics Research, PIER 10, 75–121, 1995

MODE MATCHING TECHNIQUE AS APPLIED TO OPEN GUIDED-WAVE STRUCTURES

N. Dagli

1. Introduction

The rapid and successful developments in single-mode fibers provided the motivation to develop practical integrated optical devices whose properties are optimized with respect to the fiber transmission characteristics. III-V compound semiconductors have the neces-

sary properties to realize such devices. Today, many different research laboratories are at the verge of realizing guided-wave integrated optical circuitry containing many different components like switches, modulators, power dividers, and combiners in such semiconducting materials. Computational tools for modeling and simulation are essential for successful design, optimization, and realization of such circuitry. For this purpose several different techniques exist. All of these techniques, so far, have been tried at the single component level and range from very approximate, like the effective dielectric constant (EDC) method [1], to very involved and sophisticated, like beam propagation [2,3] and finite difference methods [4]. Although sophisticated techniques are highly accurate, their computational complexity becomes a problem when the analysis has to be repeated several times to optimize a design, or needs to be extended to more complicated situations like the analysis of a circuit containing many components. On the other hand, approximate techniques are easy to use, but they are not applicable to all realistic situations and cannot provide the required accuracy. Keeping in mind that the final objective is to model and analyze an entire circuit, an equivalent circuit representation for the guided-wave optical circuitry becomes highly desirable. Such representations have been tried to solve various problems involving open guided-wave structures [5–7]. They typically originate from the so-called mode matching technique. It is the topic of this paper to describe the mode matching technique as applied to open guided-wave structures encountered in integrated optical circuits. The basic idea behind the mode matching technique is first to divide a given geometry into subsections in which the wave equation is separable. Then in each subsection the method of separation of variables is used and the field components are expressed as the product of functions of one coordinate variable only. Then one of these functions is expanded in terms of a complete set, hence mode expansions are formed for the field components. Finally these mode expansions are forced to obey the boundary conditions between the subsections. Such an approach usually results in an equivalent circuit representation for the structure under investigation. Such a representation has significant advantages. In principle an integrated optical structure can be represented as a cascade of basic building blocks. If one finds the most basic building blocks and develops a modular equivalent circuit representation for them, one can cascade the equivalent circuits representing such blocks and come up with an equivalent circuit rep-

resenting the entire structure. Then the analysis of the entire circuitry is reduced to the analysis of an equivalent circuit, which can be done and repeated many times very easily. In this paper the approach followed is based on the work of Peng and Oliner [7] and was previously reported as a journal article by the author [8]. In the description of the mode matching method, resulting in the development of an equivalent circuit, approximations valid for III-V compound semiconductors are utilized. Hence, the technique presented in this paper is especially suitable for integrated optical devices in III-V compound semiconductors. In the next section an equivalent circuit representation for waveguides based on mode matching is described. Then, the basic building blocks are identified and completely modular equivalent circuits representing them are developed. After comparing various different formulations, the technique is applied to several different structures to assess the accuracy. Finally, general conclusions are drawn.

2. Transmission Line Representation of Waveguides

In the analysis of waveguides and waveguide-related problems it is very advantageous to have a transmission line model representing the structure under examination. In such representations two distinct directions should be chosen. The first one is the axial direction defined as the z direction such that all cross sections transverse to it are identical in size and shape. Wave propagation in the waveguide is also along this direction. The other one is the transmission line direction, which is not necessarily coincident with the axial direction. One can define any other direction as the transmission line direction. In the following derivation the x direction is chosen as the transmission line or longitudinal direction and the time dependence is $e^{+j\omega t}$.

It is always possible to separate Maxwell's equations in terms of transverse and longitudinal components. Since x is defined as the longitudinal, and y and z as the transverse directions, one can write

$$\overline{E} = \overline{E}_x + \overline{E}_t \tag{1}$$

and

$$\overline{H} = \overline{H}_x + \overline{H}_t. \tag{2}$$

The desired modal representations of the transverse fields can be written as mode expansions, which are

$$\overline{E}_t(x,y,z) = \sum_i V_i(x)\overline{e}_i(y)e^{-jk_z z} \tag{3}$$

$$\overline{H}_t(x,y,z) = \sum_i I_i(x)\overline{h}_i(y)e^{-jk_z z} \tag{4}$$

where $\overline{e}_i(y)$ and $\overline{h}_i(y)$ are the transverse mode functions, which should form a complete set, and $V_i(x)$ and $I_i(x)$ are the transverse mode voltages and currents, respectively. Substituting (3) and (4) into Maxwell's equations and applying the principle of separation of variables one obtains two sets of differential equations one being a function of x and the other being functions of y only. These are

$$\frac{dV_i(x)}{dx} = -jk_{xi}Z_i I_i(x) \tag{5}$$

$$\frac{dI_i(x)}{dx} = -jk_{xi}Y_i V_i(x) \tag{6}$$

and

$$\frac{d^2\overline{e}_i}{dy^2} + (k^2 - \beta_i^2)\overline{e}_i = 0 \tag{7}$$

$$\frac{d^2\overline{h}_i}{dy^2} + (k^2 - \beta_i^2)\overline{h}_i = 0 \tag{8}$$

where

$$\beta_i^2 = k_{xi}^2 + k_z^2 \tag{9}$$

where $k_{xi}Z_i$ and $k_{xi}Y_i$ are simply the separation constants. Equations (5) and (6) simply suggest that transverse mode voltages and currents are like voltages and currents on a transmission line of characteristic impedance $Z_i = 1/Y_i$ and propagation constant k_{xi}. Equations (7) and (8) define transverse mode functions. It is possible to obtain two uncoupled sets of equations by choosing either e_{iy} or h_{iy} zero. The resulting mode sets are shown in Table I. The choices of separation constants Z_i and Y_i are also given in Table I.

Mode Set I	Mode Set II
$e_{iy} = 0$	$h_{iy} = 0$
$\frac{d^2 e_{iz}}{dy^2} + (k^2 - \beta_i^2) e_{iz} = 0$	$\frac{d^2 h_{iz}}{dy^2} + (k^2 - \beta_i^2) h_{iz} = 0$
$k_{yi}^2 = k^2 - \beta_i^2$	$k_{yi}^2 = k^2 - \beta_i^2$
$h_{iy} = -e_{iz}$	$e_{iy} = h_{iz}$
$h_{iz} = -\frac{1}{\beta_i^2} \frac{\partial^2 e_{iz}}{\partial z \partial y}$	$e_{iz} = \frac{1}{\beta_i^2} \frac{\partial^2 h_{iz}}{\partial z \partial y}$
$Y_i = \frac{\beta_i^2}{\omega \mu k_{zi}}$	$Z_i = \frac{\beta_i^2}{\omega \in k_{zi}}$
$\bar{e}_i(y) = e_{iz}(y)\bar{z}$	$\bar{e}_i(y) = e_{iy}(y)\bar{y} + e_{iz}(y)\bar{z}$
$\bar{h}_i(y) = h_{iy}(y)\bar{y} + h_{iz}(y)\bar{z}$	$\bar{h}_i(y) = h_{iz}(y)\bar{z}$

Table 1. Resulting mode sets for the choice of transmission line direction as explained in the text.

Transverse field components for either set in terms of transverse mode functions, voltages, and currents are given previously in (3) and (4). Hence, once the transverse mode functions are determined the rest of the problem requires the determination of transverse mode voltages and currents which requires the solution of transmission line equations. Therefore it is possible to represent the structure as a transmission line network and solve the problem using powerful and well-established circuit analysis techniques. These two sets are uncoupled from one another and in general both of them are needed to represent a general field distribution. In the presence of discontinuities or obstacles they may be coupled, i.e., both sets may have to be taken into account at the same time. The treatment up to this point is quite general and can be applied to any waveguide problem. This treatment is especially suitable for open rectangular dielectric waveguides. Its applications to open structures and determination of the required mode functions are explained in the next section.

3. Determination of the Mode Functions for Open Guided-Wave Structures

In Figure 1, the cross section of an open rectangular waveguide is shown. Such structures can be viewed in terms of two basic building blocks. The most suitable building blocks for structures realized in III-V compound semiconductors are the uniform regions, which are portions of multilayer slab guides and, the step discontinuities where uniform regions meet. It is advantageous to analyze and model them using the results of the previous section. Finally, models obtained can be put together to model the entire structure. Since the whole structure is composed of different multilayer slab guides choosing the slab mode functions as the mode functions of the waveguide problem is very suitable and advantageous.

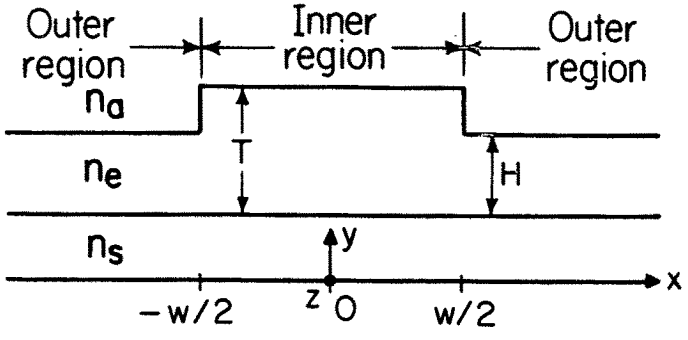

Figure 1. Cross section of an open rectangular waveguide.

At this point it could be advisable to summarize what is done so far from a slightly different point of view. A given geometry is subdivided into subregions. In each region the wave equation is separable and its solution can be written as a superposition of functions which are products of functions of one coordinate only, i.e., functions of the type $X(x)\ Y(y)\ Z(z)$. Since the direction of wave propagation in the geometry considered is in z direction and the solutions that are guided waves in z direction are the desired solutions, the functions of z should be taken as $e^{-jk_z z}$, where k_z is the unknown propagation constant. In each region functions of y are expanded using a com-

plete set as a basis. This is how equations (3) and (4) resulted. The complete set used in the mode expansions is chosen as the modes of the dielectric slab guides in that particular region. For the structure shown in Fig. 1, which is typical in integrated optics, the resulting field distributions cannot be TE or TM with respect to any given direction. In other words all six field components are present. Therefore, in the mode expansions the superposition of both TE and TM modes of the constituent regions should be used. In this example both inner and outer regions, which are uniform slab guides, have TE to y and TM to y modes. This results in two mode sets for the expansions, originating from TE and TM modes of the dielectric slabs. In Table 1 mode set I originates from the TE modes of the slabs. That is why there is no y component of the electric field associated with it. In the mode expansion superposition of such dielectric slab guide modes are used to describe the overall modes of the geometry. This situation is schematically illustrated in Fig. 2. This figure shows a TE mode of the inner slab which is just one term in the modal expansion. Clearly this mode is propagating in a certain direction and will be obliquely incident on the dielectric discontinuity as shown in the figure. A uvy coordinate system defines the slab geometry hence the slab mode is TE to y with respect to this coordinate system. The xyz coordinate system characterizes the waveguide geometry. Obviously the TE mode of the slab has E_v, H_y and H_u components. However, with respect to the waveguide coordinate system the same mode has five components, namely E_x, E_z, H_x, H_y, and H_z. Similarly a TM to y mode of the slab will have E_x, E_y, E_z, H_x, and H_z components.

For TE modes of the slab

$$\frac{H_y}{E_v} = \frac{\beta_i}{\omega\mu} \tag{10}$$

where β_i is the propagation constant of the slab mode. Since the transmission line representing this slab mode is set up along x direction its characteristic admittance, which is the same as the wave impedance of this mode in x direction, is given as

$$Y_i = \frac{V_i}{I_i} = \frac{H_y}{E_z} = \frac{H_y}{E_v \cos\theta_i} = \frac{\beta_i}{\omega\mu\cos\theta_i} \tag{11}$$

But k_{xi} and k_z, which is the propagation constant of the rib mode, are components of β_i as shown in Fig. 2, hence

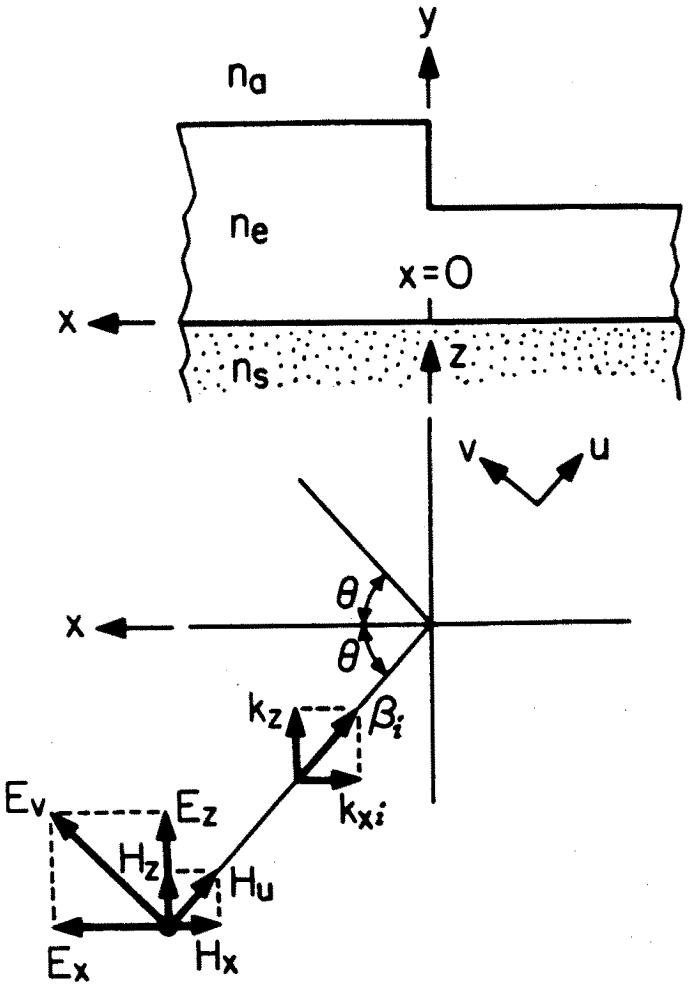

Figure 2. A TE mode of a slab guide incident obliquely on a dielectric step discontinuity.

$$k_{xi}^2 + k_z^2 = \beta_i^2 \tag{12}$$

and

$$\cos\theta_i = \frac{k_{xi}}{\beta_i} \tag{13}$$

Substituting (13) into (11) one obtains

$$Y_i = \frac{\beta_i^2}{\omega \mu k_{xi}} \tag{14}$$

One can go through a similar analysis for TM modes and find that the characteristic impedance for TM mode functions is

$$Z_i = \frac{\beta_i^2}{\omega \in k_{xi}} \tag{15}$$

The presence of E_x and H_z components of the TE mode will inevitably excite a TM wave upon scattering from the discontinuity. Therefore, TE TM mode coupling or mode conversion will occur because of the discontinuity. That is why there are not any TE or TM modes in this geometry. In other words even if only a superposition of the TE modes of the slabs is used as an input excitation to the rib waveguide, at the discontinuity they will couple into TM modes. Therefore, the resulting field distribution can only be described as a mode expansion in terms of both TE and TM modes of the slab, or in other words as a superposition of mode sets I and II in Table 1. In realistic geometries, however, the magnitude of this effect is very small and neglecting it is very well justified [11].

At this point the advantage of using the well known modes of the constituent regions in the mode expansion becomes obvious. Consider equations (3) and (4). Clearly these expansions satisfy the wave equation and the boundary conditions at any cross sectional point except at $x = \pm w/2$, since at any x other than $x = \pm w/2$ they reduce to a superposition of the modes of that particular geometry. The only points where they are not the solutions are the discontinuities where subregions meet. If the mode expansions are required to satisfy the boundary conditions at the discontinuities, a solution that satisfies boundary conditions and the Maxwell's equations everywhere is found. In principle a mode expansion can be written for the fields in each one of these subsections in terms of any complete set. But choosing the complete set used in the expansion as the complete set formed by the modes of one particular region has the advantage of satisfying the boundary conditions in that region. Therefore, one does not have to enforce boundary conditions within each subregion. As a result the analysis is simplified and the computational effort is reduced. In the next section multilayer slab guide modes are examined and transformed in a form which can be used in the treatment of the previous section.

4. Multilayer Slab Guides

In Figure 3 an asymmetric three-layer slab guide and its possible modes are shown [9]. Treatment of the multilayer slab guide problem is quite general and is given in textbooks. Therefore only a brief summary will be given. For a slab guide there are two uncoupled mode sets. The equations defining the slab mode functions are exactly the same as the equations (7) and (8) defining the mode functions of the two uncoupled mode sets defined in the previous section. Therefore it is possible and quite advantageous to use slab mode functions in the modal expansions as described earlier. The first set very much resembles the TE modes of the slab guide. Therefore it is called the TE-like mode of the structure. Similarly the second set is called the TM-like mode.

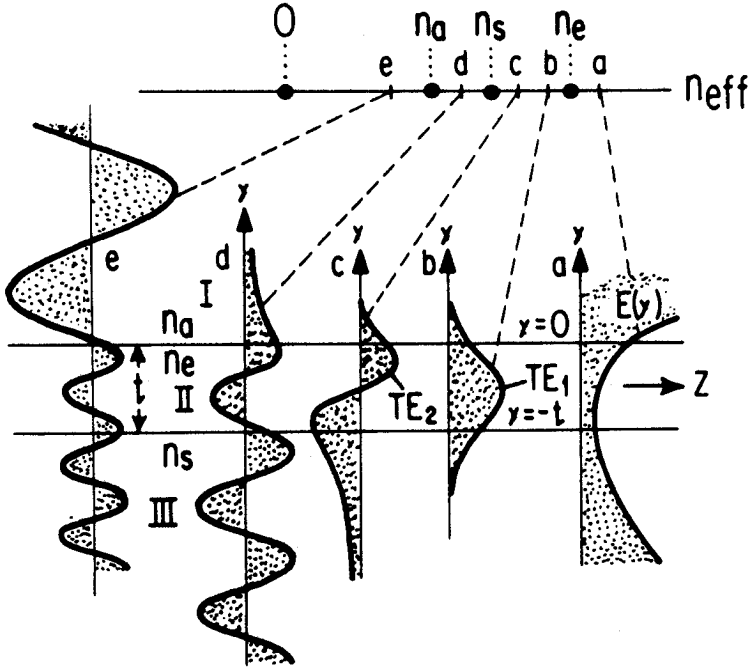

Figure 3. An asymmetric three-layer slab guide, and its possible modes.

At this point it is advantageous to normalize dimensions and indexes and work with normalized variables. According to this standard normalization, dimensions are normalized as

$$X = \frac{2\pi}{\lambda} x \sqrt{n_e^2 - n_s^2} \tag{16}$$

and effective indexes for guided slab modes as

$$b^2 = \frac{n^2 - n_s^2}{n_e^2 - n_s^2} \tag{17}$$

Indexes for continuum slab modes are normalized as

$$f^2 = \frac{n_s^2 - n^2}{n_e^2 - n_s^2} \tag{18}$$

where

$$n = \frac{\beta}{2\pi/\lambda} \tag{19}$$

β is the propagation constant of the slab mode. Then the normalized equation for the mode functions becomes

$$\frac{d^2\phi_i(Y)}{dY^2} + (b^2 - b_i^2)\phi_i(Y) = 0 \tag{20}$$

for guided modes and

$$\frac{d^2\phi(Y,f)}{dY^2} + (b^2 + f^2)\phi(Y,f)) = 0 \tag{21}$$

for the continuum modes. ϕ is either e_{iz} of mode set I or h_{iz} of mode set II. The mode functions are normalized and orthogonal to one another, i.e.,

$$\int_{-\infty}^{+\infty} \phi_i(Y)\phi_j(Y)dY = \delta_{ij} \tag{22}$$

$$\int_{-\infty}^{+\infty} \phi_i(Y)\phi(Y,f)dY = 0 \tag{23}$$

and

$$\int_{-\infty}^{+\infty} \phi(Y,f)\phi(Y,f')dY = \delta(f-f'). \tag{24}$$

Figure 3 shows the possible modes of a three-layer asymmetric slab guide for different β or b^2 values. Solutions for $\beta > k_0 n_e$ or $b^2 > 1$ shown in Fig. 3(a) are not physically realizable. For $k_o n_s < \beta < k_0 n_e$ or $0 < b^2 < 1$ guided modes exist. They are finite in number and are bound to the epilayer, i.e., they decay exponentially in the substrate and air regions as shown in Fig. 3(b) and (c). Discrete β or b^2 values are calculated as a solution of an eigenvalue equation. For $k_0 n_a < \beta < k_0 n_s$ or $0 < f^2 < a$, where $a = (n_s^2 - n_a^2)/(n_e^2 - n_s^2)$, so-called substrate or type 1 continuum modes shown in Fig. 3(d) exist. They have sinusoidal behavior in the substrate. For $0 < \beta < k_0 n_a$ or $a < f^2 < c$, where $c = n_s^2/(n_e^2 - n_s^2)$, modes have sinusoidal behavior everywhere as shown in Fig. 3(e) and are called air, or type 2 continuum modes. As the transverse variation of the continuum set gets more and more rapid, i.e., for $f^2 > c$ propagation constant values, i.e., β values become imaginary. Such modes are known as the continuum set showing cutoff behavior. Their transverse variations are similar to the set shown in Fig. 3(e) except that they are cutoff due to their rapid transverse variation. Such modes can be excited in the presence of sharp discontinuities and obstacles. For continuum modes there is no eigenvalue equation and β or f^2 values vary continuously between the indicated limits.

There is one problem, however, in using the normalized slab mode functions in the modal expansions of the waveguide problem as given by (3) and (4). Those equations are discrete summations whereas slab mode sets have both discrete and continuum components. The summations over the continuous spectra can be represented as an integration. Then the modal expansion for TE-like waveguide modes becomes

$$E_z = \sum_i^m V_i(X)\phi_i(Y) + \int_0^{+\infty} v\ (X,f)\phi(Y,f)df \tag{25}$$

$$H_y = \sum_i^m I_i(X)\phi_i(Y) + \int_0^{+\infty} i\ (X,f)\phi(Y,f)df \tag{26}$$

Although this representation is correct, it is not suitable for a modular transmission line representation. In other mode-matching techniques

given in [7], and [10] this difficulty was overcome by discretizing the continuous spectra by artificially bounding the structure with conducting planes. In this paper, however, the integration over the continuous spectrum is converted into a summation using a basis function expansion. One can expand the continuum set as

$$\phi(Y, f) = \sum_{i=1}^{\infty} \gamma_i(Y)\Psi_i(f) \tag{27}$$

where $\Psi_i(f)$ can be any orthonormal basis. Then

$$\gamma_i(Y) = \int_0^{\infty} \phi(Y, f)\Psi_i(f)df \tag{28}$$

and is called the discretized continuum set. Substituting (27) into (25) one obtains

$$E_z = \sum_{i=1}^{m} V_i(X)\phi_i(Y) + \sum_{i=1}^{\infty} C_i(X)\gamma_i(Y) \tag{29}$$

where

$$\int_0^{\infty} v(X, f)\Psi_i(f)df = C_i(X) \tag{30}$$

This representation is suitable for a transmission line representation but one still has to know the eigenvalues or normalized effective index values of discretized continuum modes. Substituting (27) into (21) and manipulating one obtains

$$\frac{d^2\Gamma}{dY^2} + b^2\Gamma + \mathbf{E}\Gamma = 0 \tag{31}$$

where

$$\Gamma^T = \{\gamma_1(Y), \gamma_2(Y), \cdots\} \tag{32}$$

and

$$E_{ij} = \int_0^{\infty} f^2\Psi_i(f)\Psi_j(f)df. \tag{33}$$

E_{ij} is the ij th element of E which is a real symmetric matrix. One can always factorize E as

$$\mathbf{E} = \mathbf{NPN}^{\mathbf{T}} \tag{34}$$

where P is the diagonal matrix with diagonal entries being the eigenvalues of E, i.e., p_i is the ith eigenvalue of E. Substituting (34) into (31) and manipulating one obtains

$$\frac{d^2\Theta}{dY^2} + (b^2\mathbf{U} + \mathbf{P})\Theta = 0 \tag{35}$$

or

$$\frac{d^2\theta_i}{dY^2} + (b^2 + p_i)\theta_i = 0 \tag{36}$$

where

$$\mathbf{N}^T\Gamma = \Theta, \tag{37}$$

and U is the identity matrix. Equation (36) is the eigenvalue equation for the transformed discretized continuum set defined by the transformation shown in (37). The eigenvalues, or normalized effective indexes of this set are p_i , i.e., the ith eigenvalue of E matrix. The guided modes and the transformed discretized continuum modes, i.e., $\{\phi_1, \phi_1, \cdots, \phi_m, \theta_1, \theta_2, \cdots, \}$, form an orthonormal mode set. Now one can utilize Table 1 and (3) and (4) and obtain a modal expansion for the field components of the waveguide modes. The resulting representation for TE-like modes is shown in Table 2. The longitudinal components, i.e., x components, are found from Maxwell's equations.

According to this formulation determination of b_z^2, $F_i(X)$, $D_i(X)$, $V_i(X)$, and $I_i(X)$ complete the solution of the problem. Since they all behave like voltages and currents on a transmission line it is possible to model uniform regions as a bunch of transmission lines. Each transmission line will correspond to a particular mode of the multilayer slab. There will be m transmission lines corresponding to the m guided modes of the slab constituting the uniform region and an infinite number of transmission lines corresponding to the continuum set of the slab guide.

$$E_y = 0 \qquad H_x = \frac{j}{\omega\mu}\sum_{i=1}^{m} V_i(X)\frac{d\phi_i(Y)}{dY} + \frac{j}{\omega\mu}\sum_{i=1}^{\infty} F_i(X)\frac{d\theta_i(Y)}{dY}$$

$$E_z = \Phi^T V + \Theta^T F$$

$$H_y = -\Phi^T I - \Theta^T D$$

$$H_z = j\sum_{i=1}^{m}\frac{\sqrt{c+b_z^2}}{c+b_i^2} I_i(X)\frac{d\phi_i(Y)}{dY} + j\sum_{i=1}^{\infty}\frac{\sqrt{c+b_z^2}}{c-p_i} D_i(X)\frac{d\theta_i(Y)}{dY}$$

$$E_x = -\omega\mu\sum_{i=1}^{m}\frac{\sqrt{c+b_z^2}}{c+b_i^2} I_i(X)\phi_i(Y) - \omega\mu\sum_{i=1}^{\infty}\frac{\sqrt{c+b_z^2}}{c-p_i} D_i(X)\theta_i(Y)$$

$$\Phi^T = \{\phi_1(Y), \phi_2(Y) \ldots \phi_m(Y)\} \quad \mathbf{V}^T = \{V_1(X), V_2(X) \ldots V_m(X)\} \quad \Gamma^T = \{\gamma_1(Y), \gamma_2(Y) \ldots\ldots\}$$

$$C^T = \{C_1(X), C_2(X) \ldots\ldots\} \quad \mathbf{F} = \mathbf{N}^T\mathbf{C} \quad \Theta = \mathbf{N}^T\Gamma$$

Table 2. Modal expansion for field components of TE-like modes of an uniform open guided-wave structure.

5. Step Discontinuities

The other constituent of the open guides are the step discontinuities where uniform regions meet. Since the modal expansions of the fields are known everywhere, step discontinuities can be modeled by forcing the fields to obey the boundary conditions at the discontinuity. The boundary conditions require the continuity of the tangential E and H fields, which in turn means

$$E_y(w/2) = \overline{E}_y(w/2) \tag{38}$$

$$E_z(w/2) = \overline{E}_z(w/2) \tag{39}$$

$$H_y(w/2) = \overline{H}_y(w/2) \tag{40}$$

$$H_z(w/2) = \overline{H}_z(w/2). \tag{41}$$

Overbars indicate the field quantities in the outer region, i.e., for $x > w/2$.

In the presence of the discontinuity, two mode sets may be coupled as described earlier. However, mode conversion due to a step discontinuity is very small and for most cases is negligible [11]. Therefore in the following discussion mode conversion will be neglected and structures will be examined under TE excitation. Writing (38)(41) explicitly using the modal expansions given in Table II one obtains

$$0 = 0 \tag{42}$$

$$\Phi^T \mathbf{V} + \Theta^T \mathbf{F} = \overline{\Phi}^T \overline{\mathbf{V}} + \overline{\Theta}^T \overline{\mathbf{F}} \tag{43}$$

$$\Phi^T \mathbf{I} + \Theta^T \mathbf{D} = \overline{\Phi}^T \overline{\mathbf{I}} + \overline{\Theta}^T \overline{\mathbf{D}} \tag{44}$$

$$\frac{d}{dY}\left\{ j \sum_{i=1}^{n} \frac{\sqrt{c + b_z^2}}{c + b_i^2} I_i(w/2)\phi_i(Y) + j \sum_{i=1}^{\infty} \frac{\sqrt{c + b_z^2}}{c - p_i} D_i(w/2)\theta_i(Y) \right\} =$$

$$\frac{d}{dY}\left\{ j \sum_{i=1}^{m} \frac{\sqrt{c + b_z^2}}{c + \overline{b}_i^2} \overline{I}_i(w/2)\overline{\phi}_i(Y) + j \sum_{i=1}^{\infty} \frac{\sqrt{c + b_z^2}}{c - \overline{p}_i} \overline{D}_i(w/2)\overline{\theta}_i(Y) \right\}. \tag{45}$$

The last equation can be approximated in the following way. Since the b^2's are normalized indexes they are between 0 and 1. For semiconducting systems suitable for optical integration $n_e \cong 3.5$ and $10^{-3} < \Delta n < 0.1$. Hence

$$17 < c = \frac{n_s^2}{n_e^2 - n_s^2} < 1700. \tag{46}$$

Therefore

$$\frac{\sqrt{c+b_z^2}}{c+b_i^2} \approx \frac{\sqrt{c+b_z^2}}{c+\bar{b}_i^2} \approx \frac{1}{\sqrt{c}} \tag{47}$$

Estimation of

$$\frac{\sqrt{c+b_z^2}}{c-p_i} \tag{48}$$

depends on the p_i values. Values of p_i depend on the basis chosen in the expansion of the continuum set but they range from very small to very large. Usually field expansions converge with only a few terms of the infinite summation. Furthermore, values of the important eigenvalues, i.e., the eigenvalues corresponding to the eigenfunctions that contribute significantly to the modal expansion, are small. This point will be shown both qualitatively and quantitatively in the next section. Hence for almost all practical structures (42) is also very close to $1/\sqrt{c}$ or when it is not almost equal to $1/\sqrt{c}$ that particular term contributes very little to the summation in (45). Then (45) reduces to

$$\frac{\mathrm{d}}{\mathrm{dY}}\{\mathbf{\Phi}^{\mathrm{T}}\mathbf{I}+\mathbf{\Theta}^{\mathrm{T}}\mathbf{D}\} = \frac{\mathrm{d}}{\mathrm{dY}}\{\overline{\mathbf{\Phi}}^{\mathrm{T}}\overline{\mathbf{I}}+\overline{\mathbf{\Theta}}^{\mathrm{T}}\overline{\mathbf{D}}\}. \tag{49}$$

But under these conditions satisfying (44) will automatically satisfy (49). Therefore, under these approximations there is no need to consider (44) and (45) individually and it is sufficient to consider (44) only.

These remaining equations can be manipulated to relate the transmission line voltages on both sides of the discontinuity in the following way. Multiplying (43) by $\overline{\Phi}$ and $\overline{\Theta}$ and (44) by Φ and Θ , integrating over Y , and using the fact that $\{\overline{\Phi}, \overline{\Theta}\}$ and $\{\Phi, \Theta\}$ form an orthonormal set one obtains

$$\overline{\mathbf{V}} = \mathbf{QV} + \mathbf{RF} \tag{50}$$

$$\overline{\mathbf{F}} = \mathbf{SV} + \mathbf{TF} \tag{51}$$

$$\mathbf{I} = \mathbf{Q}^{\mathbf{T}}\overline{\mathbf{I}} + \mathbf{S}^{\mathbf{T}}\overline{\mathbf{D}} \tag{52}$$

$$\mathbf{D} = \mathbf{R}^{\mathbf{T}}\overline{\mathbf{I}} + \mathbf{T}^{\mathbf{T}}\overline{\mathbf{D}} \tag{53}$$

where

$$\mathbf{Q} = \langle\overline{\Phi}|\Phi^T\rangle = \int_{-\infty}^{\infty} \overline{\Phi}\Phi^T dY \tag{54}$$

$$\mathbf{R} = \langle\overline{\Phi}|\Gamma^T\rangle\mathbf{N} \tag{55}$$

$$\mathbf{S} = \mathbf{M}^T\langle\overline{\Gamma}|\Phi^T\rangle \tag{56}$$

$$\mathbf{T} = \mathbf{M}^T\langle\overline{\Gamma}|\Gamma^T\rangle\mathbf{N}. \tag{57}$$

Definitions of N and M are the same except that their orders may be different depending on how many terms of the infinite summation are retained in the inner and outer regions. One can further combine (50)(57) to obtain

$$\overline{V} = GV \tag{58}$$

$$I = G^T\overline{I} \tag{59}$$

where

$$G = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{S} & \mathbf{T} \end{bmatrix} \tag{60}$$

and

$$V^T = \{\mathbf{V}^{\mathbf{T}},\ \mathbf{F}^{\mathbf{T}}\} \tag{61}$$

$$I^T = \{\mathbf{I}^{\mathbf{T}},\ \mathbf{D}^{\mathbf{T}}\} \tag{62}$$

6. Equivalent Circuit

Equations (58) and (59) immediately suggest that of a transformer. Therefore a transformer network is enough to model the discontinuity. At this point one can combine the transmission line model of the uniform regions and the transformer model of the discontinuity to model an entire structure. Figure 4 shows a rib guide and its equivalent circuit. Since the structure is symmetric only half of it has to be considered. Both uniform regions are represented as a collection of transmission lines. There are a finite number of transmission lines representing the guided modes of the uniform slab regions. In Figure 4 it is assumed that inner and outer slab regions supports k and l guided modes, respectively. The V's, I's, and Y's are the voltages, currents, and characteristic admittances of these transmission lines, respectively. The characteristic admittances are defined in Table I. In normalized format they are given as

$$Y_i = \frac{2\pi}{\lambda}\frac{\sqrt{n_e^2 - n_s^2}}{\omega\mu}\, j\, \frac{c + b_i^2}{\sqrt{b_z^2 - b_i^2}} \tag{63}$$

where b_i^2 and b_z^2 are the normalized effective indexes of the i th guided mode and the composite structure, respectively. In addition to these transmission lines there is an infinite number of transmission lines representing the continuum set. Basically these are the terms of the infinite summation and each one represents a discretized continuum mode. Again the F's, D's, and y's are the voltages, currents, and characteristic admittances of these transmission lines, respectively. The way the discretization is done allows the characteristic admittances to be defined in the same way. Using the effective index of a discretized continuum mode defined in the previous section one obtains

$$y_i = \frac{2\pi}{\lambda}\frac{\sqrt{n_e^2 - n_s^2}}{\omega\mu}\, j\, \frac{c - p_i}{\sqrt{b_z^2 + p_i}} \tag{64}$$

A set of transformers represent the discontinuity. The primary and secondary of a transformer are marked by the turns ratio of that transformer. The transformer ratios depend on the overlap integrals of the modes in the inner and outer regions and are elements of the matrices defined in (54)(57). Once the equivalent circuit is formed a transverse resonance analysis yields the propagation constant values.

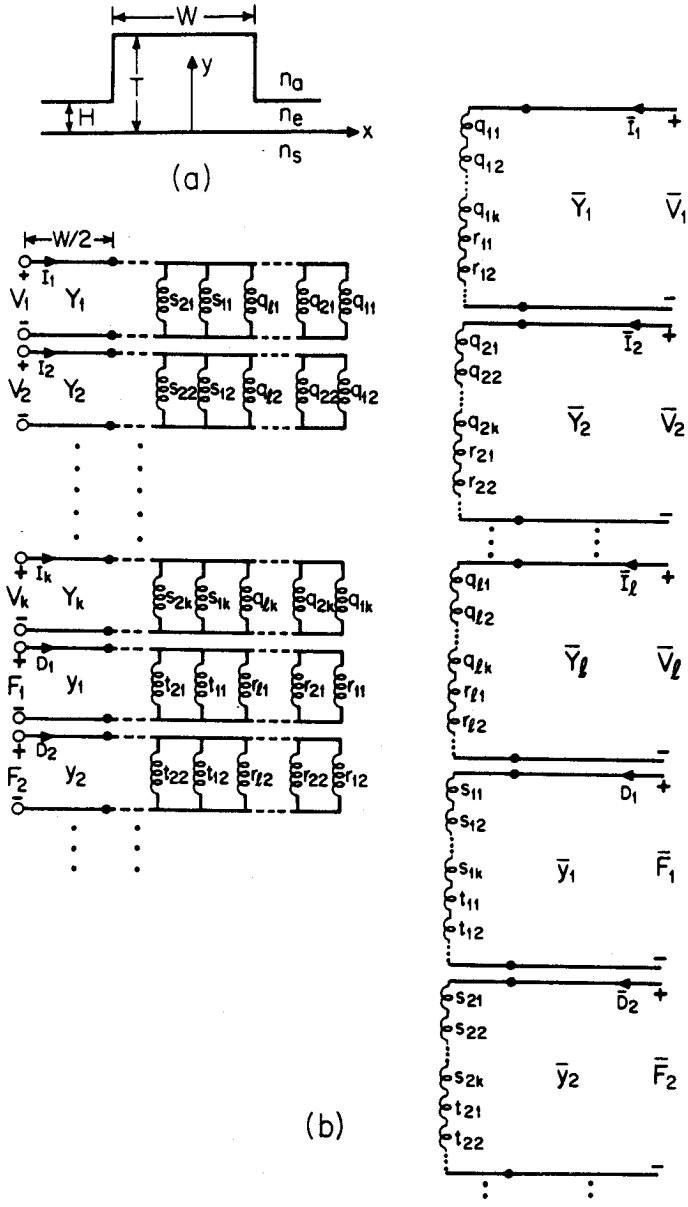

Figure 4. A rib guide and its equivalent circuit.

The equivalent circuit is physically very intuitive. A rib mode can be visualized as a superposition of the individual slab modes. These modes are obliquely incident on the discontinuity and are scattered by the discontinuity as depicted in Fig. 2. This scattering couples a particular mode to all possible modes in the inner and outer regions. The degree of coupling depends on how well the modes overlap. Then all these scattered and incident modes interfere to form the rib mode. The transformer network represents this scattering and resulting coupling.

As one can see this equivalent circuit is completely modular. Modularity makes the model very powerful. Since any open guided-wave structure can either be represented or approximated as a combination of uniform multilayer structures of various thicknesses and step discontinuities where these uniform regions meet, it is possible to model and analyze any open guided-wave structure by cascading this basic model. Therefore, the whole spectrum of integrated optical components from non ideally-shaped waveguides to non symmetrical multiple coupled structures can be analyzed. Furthermore, once the equivalent circuit parameters for the discontinuity are found, extension of the analysis from one waveguide to multiple coupled waveguides only requires the solution of a more complicated circuit which slightly increases the computational effort. Changing the widths and gaps modifies the equivalent circuit only, without the need to calculate the key circuit parameters again. This is a considerable advantage as compared with other numerical analysis techniques of comparable accuracy. For these other techniques slight changes in a given geometry or increases in the complexity of the device require the full computational effort rather than a marginal increase. Computational advantage resulting from modularity was documented in [12]. Modularity of the equivalent circuit is preserved even when there are no guided modes in a particular uniform region, because the discretization used results in a transmission line representation for all uniform regions, even those where there are no guided modes. This property also makes it possible to analyze structures where there are no guided modes in the outer regions, which cannot be analyzed with simple approximate techniques such as the EDC method. Based on this equivalent circuit a general computer program was developed to analyze single or multiple coupled waveguides with arbitrary widths and spacings. The cross-sectional profile of the waveguides can also be arbitrary.

7. Other Possible Formulations

7.1 Different Manipulations of the Boundary Conditions

The equivalent circuit presented in the previous section was obtained after a specific manipulation of equations (43) and (44). But this manipulation can be done in more than one way. To illustrate this point and the related discussion define the orthonormal sets $\{\Phi, \Theta\}$ and $\{\overline{\Phi}, \overline{\Theta}\}$ as Δ^T and $\overline{\Delta}^T$ respectively. Then (43) and (44) simply reduce to

$$E_z(x = w/2) = \Delta^T V = \overline{\Delta}^T \overline{V} \tag{65}$$

and

$$H_y(x = w/2) = \Delta^T I = \overline{\Delta}^T \overline{I} \tag{66}$$

Furthermore, in uniform regions modal voltages and currents are related to one another. For example in outer regions, which are infinitely long, $\overline{V}_i(w/2)$

$$\frac{\overline{V}_i(w/2)}{\overline{I}_i(w/2)} = \overline{Z}_i = \frac{1}{\overline{Y}_i} \tag{67}$$

where $\overline{Z}_i$ is the characteristic impedance corresponding to the i^{th} mode used in the expansion as defined in Table I. In other words

$$\overline{V} = \overline{Z}\,\overline{I} \tag{68}$$

or

$$\overline{I} = \overline{\Upsilon}\,\overline{V} \tag{69}$$

where $\overline{\Upsilon}$ and $\overline{Z}$ are diagonal matrices, whose diagonal entries are $\overline{\Upsilon}_i$ or $\overline{Z}_i$. Clearly $\overline{Z}^{-1} = \overline{\Upsilon}$. Similarly in the inner region

$$V = -\overleftarrow{Z}\,I \tag{70}$$

or

$$I = -\overleftarrow{\Upsilon}\,V. \tag{71}$$

Again $\overleftarrow{\overline{Z}}$ and $\overleftarrow{\overline{\Upsilon}}$ are diagonal matrices such that $(\overleftarrow{\overline{Z}})^{-1} = \overleftarrow{\overline{\Upsilon}}$. Each one of the diagonal entries is the input impedance of a terminated transmission line whose input is at $x = w/2$ and extends along negative x direction. In other words it is the impedance seen by an observer sitting at the inner region side of the discontinuity, i.e., at $x = w^{-}/2$, and looking at the left. The minus sign arises simply because the direction of flow of I as defined in Fig. 4 is opposite to the current flow direction used in determining the input impedance of a terminated transmission line. For example, for a single rib guide shown in Fig. 4 these terminations will either be open circuits or short circuits depending on the symmetry of the mode. So $(\overleftarrow{\overline{Z}})_i$ will be of the form $jZ_i \tan k_{xi}\frac{w}{2}$ or $-jZ_i \cot k_{xi}\frac{w}{2}$.

Equations (65) (71) can be manipulated more than one way to obtain an eigenvalue equation for the unknown propagation constant of the rib guide mode. For example, following what is done earlier one obtains

$$\overline{V} = \langle \overline{\Delta} | \Delta^T \rangle = GV \tag{72}$$

$$I = G^T \overline{I}. \tag{73}$$

But

$$I = G^T \overline{I} = G^T \overline{\Upsilon}\, \overline{V} = G^T \overline{\Upsilon} G V. \tag{74}$$

However,

$$I = -\overleftarrow{\overline{\Upsilon}} V = G^T \overline{\Upsilon} G V \tag{75}$$

which yields

$$(\overleftarrow{\overline{\Upsilon}} + G^T \overline{\Upsilon} G) V = 0 \tag{76}$$

For this equation to result in a non-trivial solution, i.e., for $V \neq 0$,

$$\det(\overleftarrow{\overline{\Upsilon}} + G^T \overline{\Upsilon} G) = 0. \tag{77}$$

The only unknown in this equation is k_z, i.e., the propagation constant of the rib mode. Therefore, equation (77) is the eigenvalue equation for the waveguide under consideration.

Equation (76) is the solution of the equivalent circuit shown in Fig. 4, and equation (77) defines the resonance conditions of this circuit. In other words at frequencies corresponding to the solution of (76) one can obtain solutions with no input excitation, which means that this particular frequency is a natural frequency of the circuit, i.e., circuit is resonant. This is the well known transverse resonance analysis. But it is possible to obtain different equations for the transverse resonance condition. Equations (65) and (66) can be manipulated to yield

$$V = \langle \Delta | \overline{\Delta}^T \rangle \overline{V} = G^T \overline{V} \tag{78}$$

$$\overline{I} = \langle \overline{\Delta} | \Delta^T \rangle I = GI \tag{79}$$

$$V = G^T \overline{V} = G^T \overline{Z}\, \overline{I} = G^T \overline{Z} G I. \tag{80}$$

But

$$V = -\overleftarrow{\overline{Z}}\, I = G^T \overline{Z} G I \tag{81}$$

or

$$(\overleftarrow{\overline{Z}} + G^T \overline{Z} G) I = 0 \tag{82}$$

which requires for a non-trivial solution

$$\det(\overleftarrow{\overline{Z}} + G^T \overline{Z} G) = 0 \tag{83}$$

Equation (82) is another solution of the equivalent circuit given in Fig. 4 and equation (83) is another transverse resonance condition which defines the eigenvalue equation for the waveguide under consideration. Similarly different manipulations yield different formulations for the transverse resonance condition or eigenvalue equation. Indeed there are eight different possible formulations. These are summarized in Table 3.

1.	$(\overline{\Upsilon} + G^T \overline{\Upsilon} G)\, V = 0$
2.	$(G\overline{\Upsilon}G^T + \overline{\Upsilon})\, \overline{V} = 0$
3.	$(G\overline{\Upsilon} + \overline{\Upsilon}G)\, V = 0$
4.	$(\overline{\Upsilon}G^T + G^T\overline{\Upsilon})\, \overline{V} = 0$
5.	$(\overline{Z} + G^T\overline{Z}G)\, I = 0$
6.	$(G\overline{Z}G^T + \overline{Z})\, \overline{I} = 0$
7.	$(G\overline{Z} + \overline{Z}G)\, I = 0$
8.	$(\overline{Z}G^T + G^T\overline{Z})\, \overline{I} = 0$

Table 3. Different formulations resulting from different manipulations of the boundary conditions at the step discontinuity.

Indeed most of these formulations are identical. This can be shown using the fact that G is a unitary matrix, i.e.,

$$G^{-1} = G^T \tag{84}$$

This simply follows from the fact that the mode sets Δ and $\overline{\Delta}$ used in the mode expansions are complete sets. It is a simple exercise to expand one complete set, say Δ , in terms of another complete set, say $\overline{\Delta}$, and show that the matrix transforming one complete set to another, G in this case, is unitary. This property can be utilized to show that formulations 1–4 given in Table 3 are equivalent to one another. As a simple example consider 1 which can be manipulated as shown below to yield 2. Multiply 1 by G from left and use $V = G^T\overline{V}$

$$G(\overleftarrow{\Upsilon} + G^T\overline{\Upsilon}G)G^T\overline{V} = (G\overleftarrow{\Upsilon}G^T + \overline{\Upsilon})\overline{V} \tag{85}$$

which is formulation 2. Similarly formulations 5–8 are all equivalent to one another because of unitarity of G. However, in practice only a finite number of terms of the complete sets Δ and $\overline{\Delta}$ are used in the expansions, hence G is not exactly unitary. Then various different formulations given in Table III are not equivalent and as a result they have different convergence properties. In other words certain formulations could be more advantageous in terms of faster convergence and accuracy. In the next section an alternate derivation of the problem is given and a physical argument is presented to identify formulations that have inherent advantages.

7.2 Variational Formulation of the Eigenvalue Equations

Starting from equations (65) and (66) one can generate an alternate formulation. Call $E_z\ (x = w/2) = \mathbf{E}$ and $H_y\ (x = w/2) = \mathbf{H}$. One can solve for V and $\overline{V}$ in terms of $\mathbf{E}$ and $\mathbf{H}$ as

$$\overline{V} = \langle \Delta | \mathbf{E} \rangle$$

$$= \left[\int_{-\infty}^{+\infty} \phi_1(Y)\mathbf{E}(Y)dY,\ \int_{-\infty}^{+\infty} \phi_2(Y)\mathbf{E}(Y)dY, \cdots \int_{-\infty}^{+\infty} \theta_1(Y)\mathbf{E}(Y)dY, \cdots \right]^T. \tag{86}$$

Using equations (70) and (71), equation (66) can be written as

$$-\Delta^T \overleftarrow{\Upsilon} V = \overline{\Delta}^T \overline{\Upsilon}\ \overline{V}. \tag{87}$$

Substituting V and $\overline{V}$ in terms of $\mathbf{E}$ and rearranging

$$\Delta^T \overleftarrow{\Upsilon} \langle \Delta | \mathbf{E} \rangle + \overline{\Delta}^T \overline{\Upsilon} \langle \overline{\Delta} | \mathbf{E} \rangle = 0. \tag{88}$$

Multiplying this equation from left by $\mathbf{E}$ and integrating over y one obtains

$$\langle \mathbf{E} | \Delta^T \rangle \overleftarrow{\Upsilon} \langle \Delta | \mathbf{E} \rangle + \langle \mathbf{E} | \overline{\Delta}^T \rangle \overline{\Upsilon} \langle \overline{\Delta} | \mathbf{E} \rangle = 0. \tag{89}$$

Since $\overleftarrow{\Upsilon}$ and $\overline{\Upsilon}$ are diagonal matrices this equation can be written as

$$\left[[\int \phi_1(Y)\mathbf{E}(Y)dY]^2, [\int \phi_2(Y)\mathbf{E}(Y)dY]^2, \cdots [\int \theta_1(Y)\mathbf{E}(Y)dY]^2 \cdots\right] \begin{bmatrix} \overleftarrow{Y}_1 \\ \overleftarrow{Y}_2 \\ \vdots \\ \overleftarrow{y}_1 \\ \vdots \end{bmatrix} +$$

$$\left[[\int \overline{\phi}_1(Y)\mathbf{E}(Y)dY]^2, [\int \overline{\phi}_2(Y)\mathbf{E}(Y)dY]^2, \cdots [\int \overline{\theta}_1(Y)\mathbf{E}(Y)dY]^2 \cdots\right] \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \vdots \\ \overline{y}_1 \\ \vdots \end{bmatrix} = 0 \tag{90}$$

This equation is an eigenvalue equation for the unknown propagation constant, k_z , of the open guide-wave structure. If the correct $\mathbf{E}(y)$ value, $\mathbf{E}_c(y)$, is substituted the correct k_z value is obtained, that is equation (90) reduces to an identity. However, in practice $\mathbf{E}_c(y)$ is not known a priori and it can only be estimated. But, k_z values obtained using this equation are stationary with respect to first order variation in $\mathbf{E}(y)$. This can be shown by calculating the variation of k_z obtained from this equation with respect to variations $\delta\mathbf{E}(y)$ of $\mathbf{E}(y)$ around $\mathbf{E}_c(y)$. This variation can be obtained by calculating the total variation of equation (90), which is

$$\int 2\delta\mathbf{E}[\Delta^T\overleftarrow{Y}\langle\Delta|\mathbf{E}\rangle + \overline{\Delta}^T\overline{Y}\langle\overline{\Delta}|\mathbf{E}\rangle]dY + [\langle\Delta^T|\mathbf{E}\rangle]^2\delta \begin{bmatrix} \overleftarrow{Y}_1 \\ \overleftarrow{Y}_2 \\ \vdots \\ \overleftarrow{y}_1 \\ \vdots \end{bmatrix}$$

$$+ [\langle\overline{\Delta}^T|\mathbf{E}\rangle]^2\delta \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \vdots \\ \overline{y}_1 \\ \vdots \end{bmatrix} = 0 \tag{91}$$

But the variations in admittances are the result of variations in k_z originating from variations in $\mathbf{E}(y)$. Hence

$$\delta \begin{bmatrix} \overleftarrow{Y}_1 \\ \overleftarrow{Y}_2 \\ \vdots \\ \overleftarrow{y}_1 \\ \vdots \end{bmatrix} = \frac{\delta}{\delta k_z} \begin{bmatrix} \overleftarrow{Y}_1 \\ \overleftarrow{Y}_2 \\ \vdots \\ \overleftarrow{y}_1 \\ \vdots \end{bmatrix} \delta k_z = F(k_z)\delta k_z$$

and

$$\delta \begin{bmatrix} \overrightarrow{Y}_1 \\ \overrightarrow{Y}_2 \\ \vdots \\ \overrightarrow{y}_1 \\ \vdots \end{bmatrix} = \frac{\delta}{\delta k_z} \begin{bmatrix} \overrightarrow{Y}_1 \\ \overrightarrow{Y}_2 \\ \vdots \\ \overrightarrow{y}_1 \\ \vdots \end{bmatrix} \delta k_z = G(k_z)\delta k_z \tag{92}$$

where elements of vectors $F(k_z)$ and $G(k_z)$ can be calculated using equations (12), (14), (63), (64) and it can be shown that they are always negative. As $\mathbf{E}(y)$ approaches $\mathbf{E}_c(y)$, i.e., as $\delta\mathbf{E}$ goes to zero the term inside the square brackets inside the integral in equation (91) goes to zero because of equation (88). Then the first term of equation (91) goes to zero even faster because it is the product of two terms each one of them approaching zero as $\delta\mathbf{E}$ approaches to zero. In other words it becomes the product of two very small terms, hence a second order term, and can be neglected compared to other first order terms. Then equation (91) reduces to

$$\{[\langle \Delta^T | \mathbf{E} \rangle]^2 F(k_z) + [\langle \overline{\Delta}^T | \mathbf{E} \rangle]^2 G(k_z)\}\delta k_z = 0. \tag{93}$$

But the term multiplying δk_z is always negative and the only way this equation can be satisfied is $\delta k_z = 0$. Therefore, the first order variations in $\mathbf{E}(y)$ will not create first order variations in k_z and the error in the propagation constant will be to second order. That means k_z values obtained using (91) are stationary with respect to the choice of a trial value for $\mathbf{E}(y)$, hence this equation represents a variational formulation for the eigenvalue equation. Therefore, using different trial

values for **E** one can obtain different approximations to the propagation constant. If one chooses a series expansion as the trial field, then Ritz procedure can be used to determine the unknown expansion coefficients. In particular, if the complete set Δ is used in the series expansion of the trial field and the Ritz procedure is applied to determine the unknown expansion coefficients, the expression given by Equation (90) can be transformed into formulation 1 shown in Table 3. This result shows that this formulation, which is obtained through a certain manipulation of the boundary conditions at the discontinuity, is the same as one would obtain from a variational expression in which the trial field at the discontinuity (which is E_z in this case) is estimated as a series in terms of the complete set used in expanding the fields in the inner region. Similarly, it can be shown that formulation 2 results from the same variational expression if the trial field is expanded in terms of the complete set used in expanding the fields in the outer region. The accuracy of the result will depend on the choice of the trial field. But since the formulation is variational, a reasonable approximation to the trial field is sufficient to get good accuracy. Then the computational efficiency depends on how many terms of the series expansion is needed to form a reasonable approximation to the trial field. In the case of a rib waveguide whose upper cladding is air, using the complete set in the inner region in approximating the trial field is more advantageous. This is because E_z at $x = w/2$ has nonzero values up to $y = T$. For $y > T$ due to a large index discontinuity between the semiconductor and air E_z value is almost zero. The same argument also applies to the guided modes of the inner slab guide. So using only the guided modes of the inner slab guide, a reasonable approximation for E_z can be formed. But this is equivalent to using only the first one or two terms of the set Δ in the expansion. Therefore, formulation 1 is expected to converge very fast. On the other hand if the set $\overline{\Delta}$ is used in forming a trial value for E_z, it will take more than the guided modes to get a reasonable approximation. This is because the guided slab modes in the outer region drop to essentially zero for $y > H$, hence part of E_z in $T > y > H$ cannot be described using these modes. Then other terms of the set $\overline{\Delta}$, i.e., terms from the discretized continuum modes, should be included which increases the required computational effort. Therefore, formulation 1 is expected to converge faster then formulation 2. Although formulations 1 and 2 are shown to originate from a variational formulation as described earlier,

the same cannot be said for formulation 3 and 4 when mode expansions do not contain infinitely many terms as in any practical calculation. Therefore, they require considerably more terms for convergence and are not advantageous. Therefore, among the admittance formulations, formulation 1 is the most advantageous one.

The same arguments apply to the other four impedance formulations. One can formulate a variational expression for the propagation constant in terms of $H_y(x = w/2)$ along the lines described earlier. Then if a trial value for $H_y(x = w/2)$ is formed as a series expansion of the terms of the set Δ and the Ritz procedure is applied, formulation 5 results. If, however, the set $\overline{\Delta}$ is used in forming a trial function then formulation 6 results. Formulations 7 and 8 cannot be obtained from the variational formulation if only a finite number of terms are used in the mode expansions. Therefore, in the light of the previous argument formulation 5 is the one that should converge faster and is the most advantageous one among the impedance formulations. Experience indicates that formulation 1 is more advantageous than formulation 5 for open guided-wave structures, hence that is the one used in formulating the equivalent circuit.

It is also interesting to see the limiting cases of this approach. Since the formulations presented are variational, different degrees of approximations can be obtained by taking into account fewer and fewer terms. As an example, consider the typical rib waveguide case where the inner and outer slabs support only one guided mode. In this case limiting the mode expansions to only one term, i.e., the mode functions corresponding to the guided modes of the slabs, an approximate analysis can be carried out. Following formulation 1 the eigenvalue equation becomes $\overleftarrow{\Upsilon} + \overline{\Upsilon}Q^2 = 0$, where $Q = \int_{-\infty}^{\infty} \phi(Y)\overline{\phi}(Y)dY$ which is the overlap integral between the guided modes of the inner and outer slabs. Writing this simple eigenvalue equation explicitly one obtains for symmetric TE modes (E_x and H_y even)

$$\left(\frac{c+\overline{b}^2}{c+b^2}\right)Q^2\left(\frac{b^2-b_z^2}{b_z^2-\overline{b}^2}\right)\tan\left\{(b^2-b_z^2)^{1/2}\frac{w}{2}\right\} = 1 \qquad (94)$$

and for anti symmetric TE modes (E_x and H_y odd)

$$-\left(\frac{c+\overline{b}^2}{c+b^2}\right)Q^2\left(\frac{b^2-b_z^2}{b_z^2-\overline{b}^2}\right)\cot\left\{(b^2-b_z^2)^{1/2}\frac{w}{2}\right\} = 1 \qquad (95)$$

where b^2's are the normalized indices defined earlier. These eigenvalue equations yield cut-off conditions which are exactly the same as the result of the EDC method. Furthermore, if $Q = 1$ the resulting eigenvalue equations are the same as those of the TM mode of a symmetric slab guide of core index n and cladding index $\overline{n}$, where n and $\overline{n}$ are the effective indices of the TE modes of the inner and outer slab regions. This, of course, is the same as EDC method. Therefore, EDC method results from a variational formulation in which the trial field at the discontinuity is chosen as the guided mode of the inner slab and the overlap integral of the inner and outer slab modes are approximated as unity. So it is obvious that EDC method will work quite accurately if the rib etching is shallow. By introducing the actual value of the overlap integral into the EDC formulation, the accuracy of the EDC method can be improved further in cases where rib etching is deep [12].

8. Application of the Equivalent Circuit to Different Structures

In this section numerical examples on rib guides of different geometries will be worked out and contributions of different terms in the equivalent circuit will be judged. The selection of rib guides is from practical considerations only. For semiconductors they are very attractive choices, because horizontal and vertical index steps are easily controllable. In the calculations it is always assumed that inner and outer regions are three-layer asymmetric slabs, i.e., there is a single epilayer on a substrate and the other cladding layer is air. The choice of a three-layer asymmetric slab is not a limitation of the method. Any number of layers can be considered, but well-known analytical expressions for normalized guided and continuum modes exist for the three-layer asymmetric slab. Furthermore, most structures can be approximated as a three-layer slab if cladding layers are thick enough. This choice determines the mode functions to be the normalized guided and continuum TE modes of a three-layer asymmetric slab guide. Laguerre polynomials weighed by $e^{-f/2}$ are chosen as the basis used in the discretization of the continuum set.

The corresponding **E** matrix up to the fifth order is given below.

$$\mathbf{E} = \begin{bmatrix} 2 & -4 & 2 & 0 & 0 \\ -4 & 14 & -16 & 6 & 0 \\ 2 & -16 & 38 & -36 & 12 \\ 0 & 6 & -36 & 74 & -64 \\ 0 & 0 & 12 & -64 & 122 \end{bmatrix} \tag{96}$$

It is a symmetric banded matrix. Eigenvalues corresponding to different orders of the E matrix are given below.

n	p_1	p_2	p_3	p_4	p_5
1	2	--	--	--	--
2	0.7889	15.211	--	--	--
3	0.42715	7.2827	46.29	--	--
4	0.26874	4.3739	24.88	98.477	--
5	0.18493	2.9409	15.989	57.212	173.67

(97)

They are the normalized effective indexes of the discretized continuum set and range from small to large for all orders. The discretized continuum set of higher orders reflect the contributions from high f values of the spectrum. This in turn means transmission lines of higher order, which correspond to higher order terms of the infinite summation in the modal expansion, represent the contributions from high f values of the continuous spectrum. But as explained in Section IV, as f increases the continuum set becomes more rapidly varying. Obviously coupling of power in the rib guide mode, which varies smoothly, to such rapidly varying modes is very small. This situation is depicted in Figs. 5 and 6. In Figure 5a guided slab mode in the inner region and a continuum mode of high f value in the outer region are shown. These two modes do no resemble one another and their overlap is very small. In Figure 6 two continuum modes of high f value are shown in the inner and outer regions. Their overlap averages out to a very small value due to their rapid variations. Hence coupling to and between continuum modes becomes very small as f values increase. This coupling is represented as the transformer ratio in the equivalent circuit. Therefore coupling being "loose" means contribution from higher order terms is

smaller and smaller, hence, addition of more terms contributes less and less and convergence occurs. In other words, from the continuous spectrum only the part around $f = 0$, i.e., substrate radiation modes of slow variation, contribute significantly to the modal expansion. Hence, it is possible to represent the structure with a small number of terms corresponding to small f values. This is especially true for formulation 1 as described earlier. In the next sections quantitative results verifying this qualitative argument will be given.

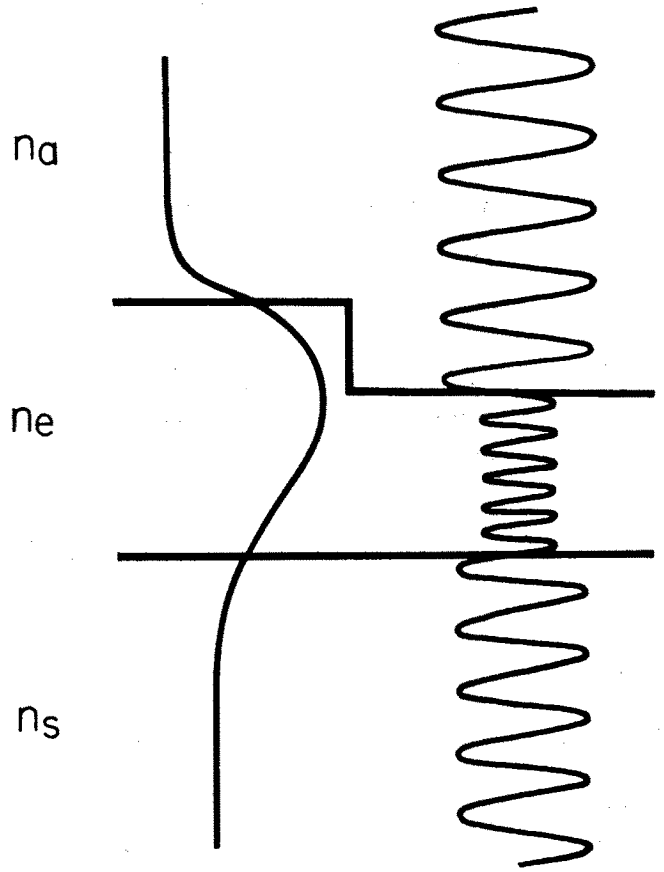

Figure 5. A guided slab mode in the inner region and a continuum mode of high f value in the outer region.

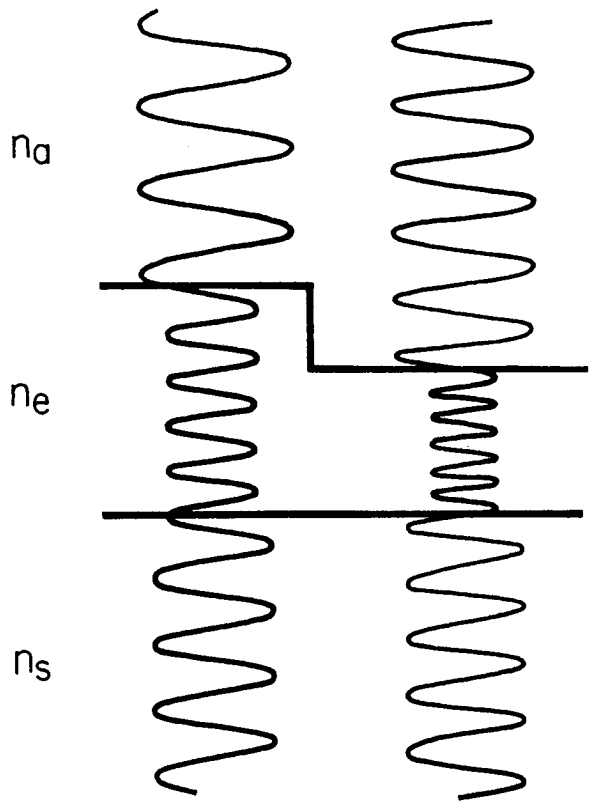

Figure 6. Two continuum modes of high f value in the inner and outer regions.

8.1 Rib Guides with No Guided Modes in the Outer Region

Rib guides with no guided modes in the outer region are used in practice to increase lateral confinement and decrease the radius of curvature of bends or increase the angle of separation of Y junctions. In the following example the rib guide shown in Fig. 7 is chosen. It is a GaAs homojunction rib guide. The index of the undoped epilayer is 3.45 at 1.15 μm wavelength. The index step between the epi and substrate is chosen as 5.25×10^{-3}. The initial epi thickness is $T = 3.85$ μm. It is etched to $H = 0.48$ μm. There are no guided modes in the outer region. Using the general computer programs developed

to analyze open guided-wave structures equivalent circuit parameters were found, circuit was solved, and normalized effective index values (b_z^2) were found for a normalized width of $W = 6.2369$, which corresponds to $W = 6$ μm. b_z^2 values as N and M (the number of discretized continuum modes in the inner and outer region, respectively) increase from 1 to 9 and are given in Table 4. As can be seen, the change in b_z^2 as the number of the terms of the infinite summation in the modal expansions increase from 1 to 9 is very little. The general trend is such that b_z^2 values increase and converge to 0.3873. The fluctuation in the last digit is due to roundoff errors and is negligible. b_z^2 values change only 0.15 percent as M and N increase from 1 to 9. This is a direct consequence of the physical argument given in the previous section. At this point it is also possible to substantiate the approximation made in satisfying the boundary conditions at the step discontinuity in Section 5. There the argument was that terms like

$$\frac{\sqrt{c + b_z^2}}{c - p_i} \tag{98}$$

can be approximated as $1/\sqrt{c}$. For the case when $M = N = 9$, the E matrix is a 9×9 matrix and p_i values keep increasing monotonically from a very small value, 0.065642, to a very large value, 725.19. In this example $c = 327$ and $b_z^2 = 0.3913$. However, a few terms are enough for convergence. If one uses the fourth term of the nine-term expansion as the highest order that needs to be considered p_4 and for that value

$$\frac{\sqrt{c + b_z^2}}{c - p_4} = \frac{\sqrt{327 + 0.3913}}{327 - 17.04} = 0.0583 \tag{99}$$

whereas

$$\frac{1}{\sqrt{c}} = 0.0553 \tag{100}$$

Hence approximating (98) as $1/\sqrt{c}$ is still valid. For higher order terms, the approximation breaks down, but their contribution is negligible as explained and quantitatively shown before. Hence the basic approximation in the derivation of the model is quantitatively justified.

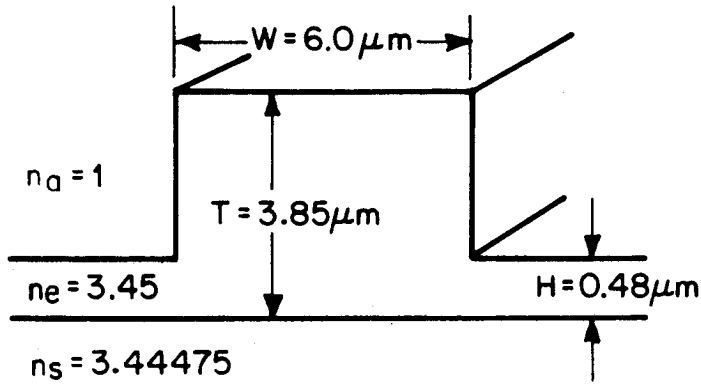

Figure 7. A GaAs homojunction rib guide with no guided modes in the outer region.

N	0	1	2	3	4	5	6	7	8	9
M	1	1	2	3	4	5	6	7	8	9
b_z^2	0.3867	0.3866	0.3874	0.3871	0.3874	0.3875	0.3874	0.3874	0.3873	0.3873

Table 4. Variation of the normalized effective index of the rib guide in Figure 7 as a function of the number of discretized continuum modes in the inner and outer regions.

8.2 Rib Guides with Guided Modes in the Inner and Outer Regions

Rib guides with guided modes in the inner and outer regions are used in practice for couplers where weak confinement is desired to obtain short coupling lengths. In the following example the rib guide shown in Fig. 8 is used. It is the same guide used in the previous example except that the epilayer in the outer region is etched to 3.37 μm.

Hence the outer region supports a guided mode. Normalized effective index values b_z^2 as a function of increasing N and M are shown in Table 5. Again the same general features are observed. Convergence is very rapid and the result obtained with up to the nine discretized continuum modes taken into account in each region, is within 0.5 percent of the result obtained when the continuous spectra is neglected.

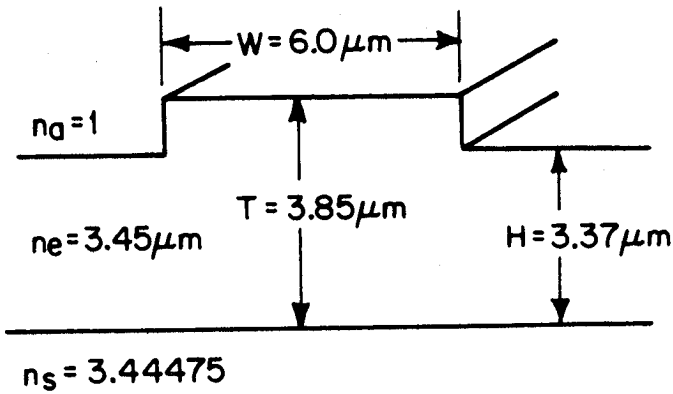

Figure 8. A GaAs homojunction guide with guided modes both in the inner and outer regions.

N	0	1	2	3	4	5	6	7	8	9
M	0	1	2	3	4	5	6	7	8	9
b_z^2	0.5767	0.5723	0.5735	0.5733	0.5735	0.5735	0.5735	0.5735	0.5735	0.5735

Table 5. Variation of the normalized effective index of the rib guide in Figure 8 as a function of the number of discretized continuum modes in the inner and outer regions.

The last two sections illustrate that for practical structures, even in the extremes of the waveguide geometry convergence is obtained with limited computational effort. In between the extremes, especially close to the cutoff of the guided mode in the outer region, more terms of the circuit may be needed for convergence but usually less than five terms is enough [12]. These results are representative of the general properties of the technique and other results for a variety of guides and index steps demonstrate the same behavior [14]. These results also indicate that the approximations made in (47) and (48) are well justified for the structures of practical interest. Rapid convergence of the technique makes it possible to simplify the equivalent circuit considerably, so that it can be solved by hand using a hand calculator [15].

The other obvious question is the accuracy. Accuracy is checked by comparing the results of the present analysis with the results of the other numerical and analytical techniques that exist in the literature for several different waveguides. This is done in the next section.

9. Comparison with other Techniques

9.1 Rectangular Fiber

The rectangular fiber shown in Figure 9 is analyzed by Goell [16] extensively. The only practical semiconductor waveguide that can be approximated as a rectangular fiber is the buried heterostructure guide, which is not easy to fabricate [17]. Fabrication of this structure either requires overgrowth [17] or mass transport [18]. But since Goell's analysis is verified by independent researchers using different techniques it is a good choice to check the accuracy of the present method.

In the first example the index step between the core and the outer region is very small. Actual n_e and n_s values do not matter as long as $c > 10$. In the numerical calculations $n_e = 1.00142$ and $n_s = 1.0$, which corresponds to $c = 350$ and $a = 0$, since slabs are symmetric. Fig. 10 shows the dispersion diagram for the E_{11}^x mode for two different aspect ratios, $W = T$ and $W = 2T$. In this figure dimensions are normalized slightly differently to be consistent with Goell's definitions, which is

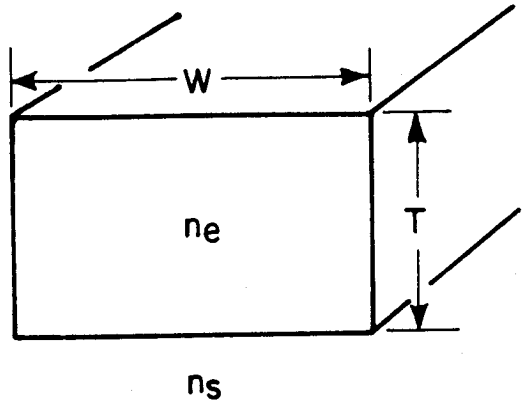

Figure 9. A rectangular fiber.

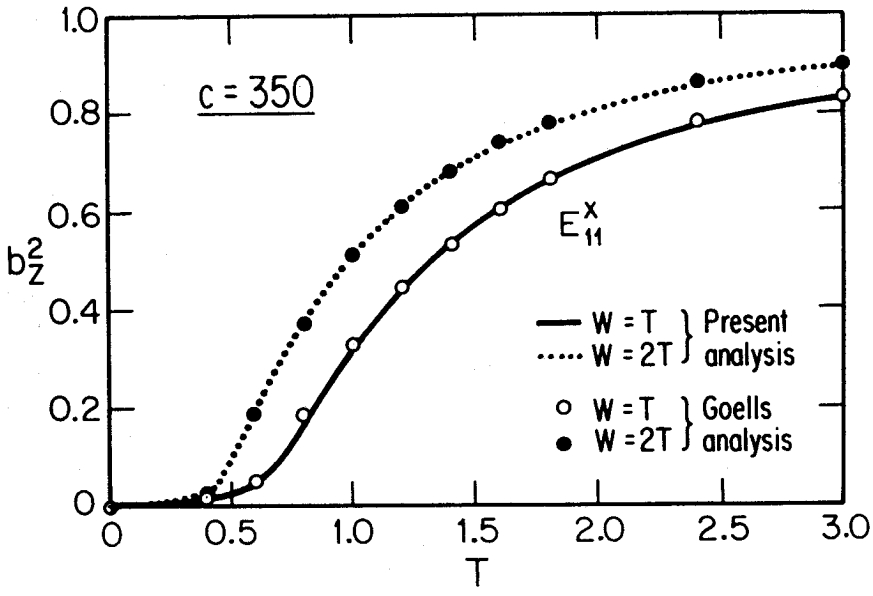

Figure 10. Dispersion diagram of the fundamental mode of the rectangular fiber for $c = 350$.

$$T(W) = \frac{2}{\lambda} t(w) \sqrt{n_e^2 - n_s^2}. \tag{101}$$

This definition lacks a factor of π with respect to the normalization defined previously and used in this chapter. Continuous curves show the results of the present analysis and points are the results of the Goell's analysis. Agreement for both aspect ratios over the entire spectrum is excellent and results agree to better than 1 percent. For the aspect ratio of two, i.e., $W = 2T$ for $T > 1$ using only one discretized continuum mode in the outer region and the guided mode(s) in the inner region gives an accuracy better than 3 percent. This simplifies the equivalent circuit and numerical calculations considerably. The same observation is true for larger aspect ratios starting even at smaller T values. For $T < 1$, however, one has to take progressively more terms for convergence and accuracy. For the unity aspect ratio, i.e., $W = T$, at least two terms have to be taken into account over the entire spectrum. For $T < 0.5$ up to ten terms had to be taken into account to get convergence. This is because for small aspect ratios and T values, fields spread outside the core more and more, hence, the detailed representation of the fields in the outer region becomes more important, which in turn requires more terms in the expansions. Using more terms, however, gives more accurate results. In Figure 11 the dispersion diagrams for the E_{11}^x mode of a rectangular fiber with a large index step for two different aspect ratios are shown. For this case $c = 0.8$, which corresponds to $n_e = 1.5$ for $n_s = 1.0$. This is a very small c value and for such small c values approximations made in Section V in satisfying the boundary conditions become too approximate. However, it is interesting to see the results of the method for this case. In Figure 11 continuous lines are the result of the present analysis and the points are the results of the Goell's analysis. For $W = 2T$ the two results agree very well for $T > 1$. For $W = T$, however, very good accuracy is obtained for $T > 2$. But that requires at least five terms. For small T values results of the present analysis are higher than Goell's analysis. In this range in order to obtain convergence approximately ten terms have to be taken into account. Agreement for $W = 2T$ is better. These results indicate that, even when the basic approximation breaks down, it is possible to get sufficiently accurate results for large aspect ratios using a large number of terms. This is because unless $c > p_i$ the mode expansion is in terms of the continuum set showing cutoff behavior, which cannot express a guided mode. But since $c = 0.8$, only when a

high number of terms are considered do the first couple of eigenvalues become less than 0.8 and does the mode expansion become proper. But in order to have a sufficient number of terms such as $p_i < c$ one has to go to very high orders. Even for $N = 5$ only $p_1 < 0.8$, as shown in (97). Therefore, when a high number of terms, like ten, is considered sufficient number of terms will have $p_i << c$, hence, the basic assumption will be satisfied and the results will become accurate. This, of course, increases computational complexity and makes the method comparable to the complexities of other numerical techniques, but modularity is still preserved and remains as an advantage. At this point it should be stressed that structures with small c values and aspect ratios are seldom found in practice. Even for a buried heterostructure guide, typical c values and aspect ratios are larger than 7 or 10, [17] hence, the basic assumptions are valid in the vast majority of practical cases.

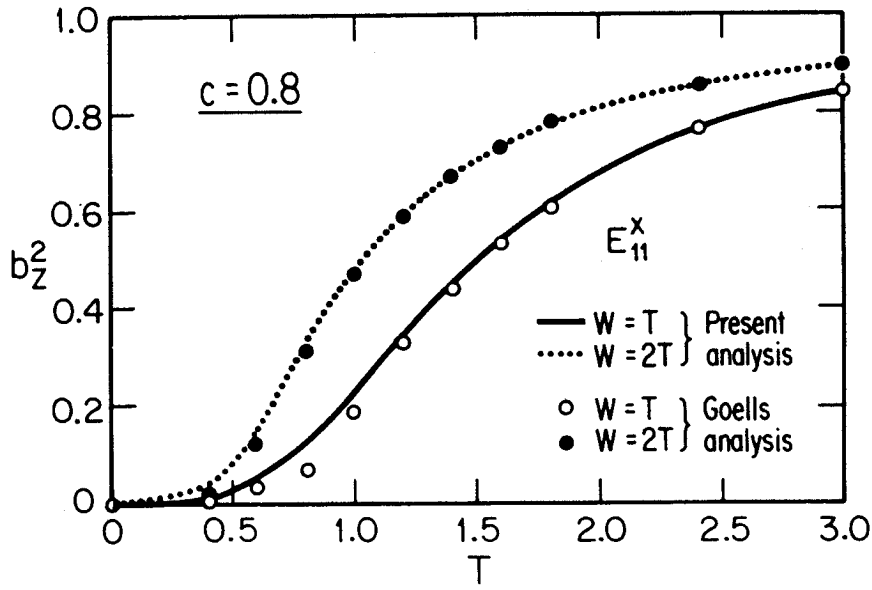

Figure 11. Dispersion diagram of the fundamental mode of the rectangular fiber for $c = 0.8$.

9.2 Rib Guides with Sloped Rib Sides

In practice rib guides fabricated using chemical etching techniques tend to have sloped rib sides. One has to come up with a technique to analyze rib guides with sloped rib sides. This is done in this work by approximating the sloped side as a staircase function [19]. The step-like sidewall approximated this way can in turn be modeled cascading the basic building block described in this work.

Pelossi *et al.* [20] analyzed sloped side rib guides using a finite element method (FEM) analysis. In the analysis they used triangular meshes to fit the geometry better. Their analysis was applied to sputtered glass films on glass substrates and their theoretical results were verified experimentally. The structures that are analyzed in [20] were also analyzed using the present method, in which the sloped side is approximated as a staircase function. Figures 12(a) and (b) show the two structures that were analyzed. The guide in Fig. 12(a) has a sidewall angle of 21° and the outer region supports a guided mode. For this case five steps were found to be adequate for the convergence of the normalized index b_z^2. Normalized index values for the E_{11} mode calculated for different W values using only guided modes and three discretized continuum modes in the inner and outer regions are shown in Table 6 together with the results of the FEM analysis of [20]. Agreement is excellent demonstrating the accuracy of the staircase approximation to the sloped side. Another observation is that, as the number of discretized continuum modes increases from 0 to 3, the normalized index changes negligibly, indicating that the effect of continuous spectra is negligible for well-guided modes, as demonstrated in Section 8.2. The structure in Fig. 12(b) has no guided modes in the outer region and has a sidewall angle of 32°. For this case seven steps were found to be adequate for convergence. Results of the present analysis and FEM analysis for the E_{11} mode again are shown in Table 7. The agreement is excellent, and normalized index values change negligibly as the number of discretized continuum modes are increased. These comparisons demonstrate the accuracy of approximating the sloped side as a staircase function. This approximation coupled with the modularity and accuracy of the present method of analysis, makes it a very powerful tool to analyze and design open waveguides of arbitrary profiles. Modularity even when there are no guided modes in the outer region is the key factor. Modularity of other mode-matching techniques that come up with equivalent circuits [11] breaks when there are no guided modes

in the outer regions, hence, they cannot be used to analyze structures like the one in Fig. 12(b).

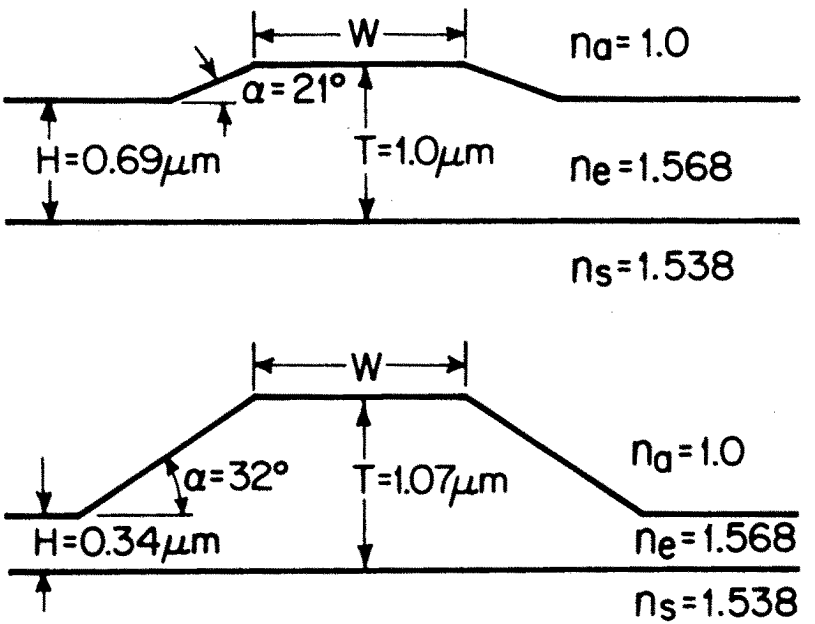

Figure 12. The sloped side rib guides analyzed for comparison with FEM analysis.

W(μm)	b_z^2(N = M = 0)	b_z^2(N = M = 3)	b_z^2(*FEM* result)
4.27	0.4612	0.4610	0.4642
6.61	0.4748	0.4747	0.4716
7.88	0.4786	0.4784	0.4790
10.49	0.4789	0.4786	0.4791

Table 6. Variation of the normalized effective index of the rib guide in Figure 12(a) as a function of the number of discretized continuum modes in the inner and outer regions.

W(μm)	b_z^2(N = 0 M = 1)	b_z^2(N = M = 3)	b_z^2(*FEM* result)
2.66	0.4587	0.4583	0.4531
4.88	0.4912	0.4910	0.5031
7.44	0.5185	0.5183	0.5124
9.49	0.5210	0.5206	0.5180

Table 7. Normalized index values for the E_{11} mode of the structure in Figure 12(b) as calculated by the present method and FEM analysis.

In addition to these comparisons experiments were performed on GaAs rib guides, directional couplers, and three-guide couplers [21]. The parameters describing these structures were experimentally determined, and using these parameters structures were simulated using the present method of analysis. The number of modes supported by waveguides of different widths and the transfer lengths of directional and three-guide couplers were determined both experimentally and through simulations. Agreement between the theoretical and experimental results were very good, once again verifying the accuracy of the technique. The accuracy and limitations of the commonly-used approximate techniques for the analysis of coupled structures were also assessed [21]. Using this technique universal-design curves for the design of rib guides were also generated [12]. Simple approximate formulas for rib guide design and the device parameter ranges over which these approximate formulas are valid were also given in [12].

10. Conclusion

In this paper the mode matching technique as applied to open guided-wave structures is studied. It is shown that this technique can result in a modular equivalent circuit representation to model and analyze open guided-wave structures. The equivalent circuit developed takes into account both the guided and continuous spectra. Using a

basis function expansion the continuous spectra was discretized and transformed into a form which is suitable for a modular equivalent circuit representation. In the development of the method, approximations which are valid for integrated optics in III-V semiconductors were made. The validity of these approximations was demonstrated both qualitatively based on a physical argument and quantitatively based on numerical simulations on rib guides. It is also shown that the mode matching technique can yield many different formulations. Although in principle all these formulations are equivalent, in practice they all have different convergence properties. Using a variational formulation, it is shown that the formulations that will yield faster convergence can be chosen based on physical arguments. This formulation also shows that the equivalent circuit representation originated from a variational analysis and has the potential to yield very accurate results with very limited number of terms. In order to assess the accuracy of the technique, comparisons were made with the results of other methods of analysis for the rectangular fiber and rib guides with sloped rib sides. These comparisons indicate that present technique yields accurate results with limited computational effort. Modularity becomes especially useful to analyze structures with arbitrary cross-sectional profiles or when the analysis has to be repeated or extended to more complicated situations.

References

1. Knox, M. R., and P. P. Toulios, *Proceedings of the Symposium on Submillimeter Waves,* New York, Polytechnic, p. 497, 1970.
2. Feit, M. D., and J. A. Fleck, Jr., "Light propagation in graded index optical fibers," *Appl. Opt.,* Vol. 17, No. 24, 3990–3998, Sept. 15, 1978.
3. Chung, Y., and N. Dagli, "An assessment of finite difference beam propagation method," *IEEE J. Quantum Electron.,* Vol. QE-26, No. 8, 1335–1339, Aug. 1990.

4. Robertson, M. J., S. Ritchie, and P. Dayan, "Semiconductor waveguides: analysis of optical propagation in single rib structures and directional couplers," *IEE Proc.*, Vol. 132, pt. J., No. 6, December 1985.
5. Rozzi, T. E., and G. H. Intveld, "Field and network analysis of interacting step discontinuities in planar dielectric waveguides," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-27, 303–309, Apr. 1979.
6. Koshiba, M., and M. Suzuki, "Vectoral wave analysis of dielectric waveguides for optical integrated circuits using equivalent network approach," *J. Lightwave Technol.*, Vol. LT-4, 656–664, June 1986.
7. Peng, S. T., and A. A. Oliner, "Guidance and leakage properties of a class of open dielectric waveguides: Part I," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, 843–855, Sept. 1981.
8. Dagli, N., "Equivalent circuit representation of open guided-wave structure," *IEEE J. Quantum Electron.*, Vol. 26, 98–108, Jan. 1990.
9. Yariv, A., *Introduction to Optical Electronics,* p. 356, 2nd edition, Holt, Rinehart and Winston, 1976.
10. Mittra, R., Y. L. Hou, and V. Jamnejad, "Analysis of open dielectric waveguides using mode matching technique and variational methods," *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-28, 36–43, Jan. 1980.
11. Koshiba, M., and M. Suzuki, "Vectoral wave analysis of optical waveguides with rectangular cross section using equivalent network approach," *Electron. Lett.*, Vol. 18, 473–475, Sept. 15, 1985.
12. Dagli, N., and C. G. Fonstad, "Universal design curves for rib waveguides," *J. Lightwave Technol.*, Vol. 6, 1136–1145, "Special Issue on Integrated Optics," June 1988.
13. Collin, R. E., *Field Theory of Guided Waves,* Chapter 8, McGraw-Hill, 1960.
14. Dagli, N., and C. G. Fonstad, "Microwave equivalent circuit representation of rectangular dielectric waveguides," *Appl. Phys. Lett.*, Vol. 49, No. 6, 308–310, Aug. 11, 1986.
15. Dagli, N., and C. G. Fonstad, "Analysis of rib dielectric waveguides," *IEEE J. Quantum Electron.*, Vol. QE-21, 315–321, Apr. 1985.

16. Goell, J. E., "A circular harmonic computer analysis of rectangular dielectric waveguides," *Bell Syst. Tech. J.*, Vol. 48, No. 9, 2133–2160, Sept. 1969.
17. Johnson, L. M., "Low loss GaInAsP buried heterostructure optical waveguide branches and bends," *Appl. Phys. Lett.*, Vol. 44, 278–280, Feb. 1, 1984.
18. Liau, Z. L., and J. N. Walpole, "A novel technique for GaInAsP/InP buried heterostructure laser fabrication," *Appl. Phys. Lett.*, Vol. 40, 568–570, Apr. 1, 1982.
19. Dagli, N., and C. G. Fonstad, "Analysis of rib waveguides with sloped rib sides," *Appl. Phys. Lett.*, Vol. 46, No. 6, 529–531, Mar. 15, 1985.
20. Pelossi, P. M., P. Vandenbulcke, C. D. Wilkinson, and R. M. DeLaRue, "Propagation characteristics of trapezoidal cross section ridge optical waveguides: an experimental and theoretical investigation," *Appl. Opt.*, Vol. 17, No. 8, 1187–1193, Apr. 15, 1978.
21. Dagli, N., and C. G. Fonstad, "Theoretical and experimental study of analysis and modeling of integrated optical components," *IEEE J. Quantum Electron.*, Vol. 24, 2215–2228, Nov. 1988.