Helmholtz beam propagation in rib waveguides and couplers by iterative Lanczos reduction

R. P. Ratowsky, J. A. Fleck, Jr., and M. D. Feit

Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94550

Received June 10, 1991; accepted August 26, 1991

The solution of the Helmholtz equation requires the application of an exponentiated square root operator to an initial field. This operation is greatly facilitated by the introduction of a representation in which the above-mentioned operator is diagonal. The Lanczos method permits this diagonalization to be performed in a low-dimensional space and the propagation to be carried out to arbitrary order. An iteration scheme to be carried out in conjunction with the Lanczos method is also described. This scheme permits, in principle, the bound mode of a monomode waveguide to be calculated to machine accuracy. Results for some well-known test examples of rib waveguides are presented.

1. INTRODUCTION

A new method for solving the scalar Helmholtz equation based on iterative Lanczos reduction was recently proposed. The accuracy of the method was established by a sixth-order calculation of the field resulting from the injection of a large number of modes (26) into a quadratic waveguide. Spectral analysis of the propagating field confirmed the continuing presence of all the injected modes and verified that the accompanying propagation constants were the correct Helmholtz eigenvalues to high precision. These results established both the correctness and the accuracy of the Lanczos solution method for the Helmholtz equation and suggest that the method may be suitable for solving a wide range of problems for which paraxial or wide-angle-paraxial methods may be inadequate. In the present paper we further explore the method by applying it to rib waveguide structures.

The growing importance of semiconductor rib waveguides has created the need for accurate numerical modeling capabilities. At the same time, the structure and abrupt changes in refractive index that characterize these devices present a challenge to the robustness and accuracy of any numerical method. A variety of techniques have so far been applied to the analysis of rib waveguides. These include scalar finite difference methods [FD(1), FD(2)],5 vector finite element methods,5 mode matching,7 the Rayleigh Ritz method,8 semivectorial finite-difference methods [FD(3), FD(4)],9 the matrix beam propagation method,9 the straightforward beam propagation method (BPM),10 and a variational method based on simple functional forms for the electric field.11 Both the beam propagation methods listed above were based on a paraxial wave equation.

The objective of all the studies cited above was the determination of guided mode fields and corresponding propagation constants. None of the methods listed, however, would be suitable for describing transient field effects. The BPM computations,10 for example, were used to determine paraxial mode eigenvalues, which in turn were corrected by a square-root transformation to determine Helmholtz eigenvalues. But because the structures were strongly guiding, the transient paraxial solutions could not be expected to describe the contribution of the radiation modes to the field correctly. Iterative Lanczos reduction solutions to the Helmholtz equations, on the other hand, can be expected to describe accurately both radiation and guided modes.

2. WAVE EQUATIONS

We shall be concerned with a scalar electric field \( \Psi(x, y, z) \), which satisfies the scalar Helmholtz equation

\[
\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} + \frac{n^2(x, y) \alpha^2}{c^2} \Psi = 0. \tag{1}
\]

For the purposes of the present study the refractive-index distribution \( n(x, y) \) is assumed to depend only on the transverse coordinates \( x \) and \( y \). It is convenient, although by no means necessary for the application of the Lanczos method, to factor a carrier wave from the field. Thus we write \( \Psi(x, y, z) = \exp(-ikz)\psi(x, y, z) \), where \( k = \omega n_0/c \) and \( n_0 \) is a reference refractive index. Substitution of this expression into Eq. (1) gives the following equation for \( \psi \):

\[
-\frac{1}{2k} \frac{\partial^2 \psi}{\partial z^2} + \frac{i}{2k} \frac{\partial \psi}{\partial z} = -\frac{1}{2k} \nabla_x^2 \psi + \frac{k}{2} \left[ \frac{n^2(x, y)}{n_0^2} - 1 \right] \psi. \tag{2}
\]

Neglect of the second \( z \) derivative gives the paraxial wave equation

\[
\frac{i}{2k} \frac{\partial \psi}{\partial z} = \frac{1}{2k} \nabla_x^2 \psi + \frac{k}{2} \left[ \frac{n^2(x, y)}{n_0^2} - 1 \right] \psi. \tag{3}
\]

3. SOLUTION IN TERMS OF MODES

General solutions to Eqs. (2) and (3) can be developed by expansion in terms of a complete set of modes appropriate to each equation. If one includes \( z \) dependence, these modes can be expressed in the form

\[
\psi(x, y, z) = \phi_n(x, y) \exp(-i\beta_n z), \tag{4a}
\]

\[
\psi(x, y, z) = \psi_n(x, y) \exp(-i\beta_n z), \tag{4b}
\]
where the prime indicates a solution of the paraxial wave equation. Substituting expressions (4a) and (4b) into Eqs. (2) and (3), respectively, one finds that

\[ \psi_n(x, y) = \psi_n(x, y), \]  
\[ \beta_n = -i/(1 + 2\beta_n/k)1^{2\gamma}. \]  

(5a)  
(5b)

These results follow from the fact that Eqs. (2) and (3) can be written as

\[ \frac{1}{2k} \frac{\partial^2 \psi}{\partial z^2} + \frac{i}{2k} \frac{\partial \psi}{\partial z} = \mathbf{H} \psi, \]  
\[ \frac{i}{2k} \frac{\partial \psi}{\partial z} = \mathbf{H} \psi, \]  

(6a)  
(6b)

where the operator \( \mathbf{H} \) is defined by

\[ \mathbf{H} = \frac{1}{2k} \nabla^2 + \frac{k}{2} \left[ n^2(x, y) - 1 \right] \]  

(7)

and the eigenvalues of \( \mathbf{H} \) are \( \beta_n \). In Eq. (5b) the sign of the square root has been chosen to ensure that the wave moves in the positive \( z \) direction.

Making use of Eqs. (5a) and (5b), one can express solutions to Eqs. (2) and (3) in the following form:

\[ \psi(x, y, z) = \sum_{n=0}^{N} A_n \psi_n(x, y) \exp\{i \left[ k \left( (1 + 2\beta_n/k)1^{2\gamma} \right) z \right] \}, \]  
\[ \psi'(x, y, z) = \sum_{n=0}^{N} A_n \psi_n(x, y) \exp\{-i \beta_n z\}, \]  

(8a)  
(8b)

where \( \psi'(x, y, 0) = \psi(x, y, 0) \).

4. NUMERICAL SOLUTION BY MATRIX DIAGNOLIZATION AND ITS RELATION TO THE SOLUTION IN TERMS OF MODES

One can, in principle, determine both the modes and the solutions that correspond to Eqs. (8) by matrix diagonalization.  

One way to accomplish this is to introduce a finite rectangular \( N \times N \) grid of side \( L \) and a finite basis of plane-wave functions

\[ u_{mn}(x, y) = \frac{1}{L} \exp \left[ \frac{2\pi i}{L} \left( mn + ny \right) \right], \]
\[ \frac{N}{2} < m \leq \frac{N}{2}, \quad \frac{N}{2} < n \leq \frac{N}{2}. \]  

(9)

In this representation the field can be expressed as the finite Fourier series

\[ \psi(x, y, z) = \sum_{m=-N/2+1}^{N/2} \sum_{n=-N/2+1}^{N/2} B_{mn}(z) \exp \left[ \frac{2\pi i}{L} \left( mn + ny \right) \right], \]  

(10)

and the set of all values \( B_{mn}(z) \) can be regarded as an \( N^2 \)-dimensional vector \( \psi(z) \). If we form an \( N^2 \times N^2 \) matrix \( \mathbf{H} \) from the elements \( \langle mn|\mathbf{H}|m'n' \rangle \), defined in the usual complex Hilbert space sense, Eqs. (6a) and (6b) can be written as

\[ -\frac{1}{2k} \frac{\partial^2 \psi}{\partial z^2} + \frac{i}{2k} \frac{\partial \psi}{\partial z} = \mathbf{H} \psi, \]  
\[ \frac{i}{2k} \frac{\partial \psi}{\partial z} = \mathbf{H} \psi. \]  

(11a)  
(11b)

If Eqs. (11a) and (11b) are multiplied through by the unitary matrix \( \mathbf{U} \), which converts \( \mathbf{H} \) to diagonal form, the result can be written as

\[ -\frac{1}{2k} \frac{\partial^2 \psi}{\partial z^2} + \frac{i}{2k} \frac{\partial \psi}{\partial z} = (\mathbf{H} \mathbf{U}^{-1}) \psi, \]  
\[ \frac{i}{2k} \frac{\partial \psi}{\partial z} = (\mathbf{H} \mathbf{U}^{-1}) \psi. \]  

(12a)  
(12b)

The solutions to Eqs. (12a) and (12b) can be expressed formally as

\[ \psi(z) = \exp[i k \left( (1 + 2\beta/k)1^{2\gamma} \right) z] \psi(0), \]  
\[ \psi(z) = \exp(-i \beta z) \psi(0). \]  

(13a)  
(13b)

Solutions (14a) and (14b), which correspond to solutions (8a) and (8b) in terms of modes, are expressed in the momentum or plane-wave representation, i.e., coefficients of the type \( B_{mn} \) form the components of the vector. To express the vector components as function values on a grid it is necessary to apply an inverse discrete Fourier transform, \( \mathcal{F}^{-1} \), with the help of a fast Fourier transform to the components of \( \psi(z) \). The eigenvectors of \( \mathbf{H} \), which are equal to the columns of \( \mathbf{U} \), correspond directly to modes (4a) and (4b). The function values of the modes on the grid can be likewise formed by taking inverse discrete Fourier transforms of the eigenvectors of \( \mathbf{H} \).

5. SOLUTION BY ITERATIVE LANCZOS REDUCTION

While it is possible in principle, the above technique is clearly impractical for any but the sparsest grids. If, however, we are willing to advance the solution over a limited propagation step \( \Delta z \) rather than over an arbitrary propagation distance \( z \), it becomes possible to carry out the same procedure using a much smaller basis. The vectors \( \psi(0), \mathbf{H}\psi(0), \ldots, \mathbf{H}^n \psi(0) \), also known as the Krylov vectors, where \( \mathbf{H} \) is evaluated in representation (10), can be used to form such a basis. This basis is suggested by the finite Taylor series expansion of the solution to Eq. (11b) in terms of an evolution operator

\[ \psi(\Delta z) = \exp(-i \Delta z \mathbf{H}) \psi(0) = \sum_{n=0}^{N} \frac{1}{n!} (-i \Delta z \mathbf{H})^n \psi(0) + O((\Delta z)^{N+1}). \]  

(15)

The Krylov vectors are not orthonormal, but the following recursion procedure, which is due to Lanczos, can be used to form from them an orthonormal set:

\[ q_0 = \psi(0), \]  
\[ \mathbf{H} q_n = \alpha_n q_n + \beta_n q_{n-1}, \]  
\[ \mathbf{H} q_n = \beta_{n+1} q_{n+1} + \alpha_n q_n + \beta_n q_{n-1}, \]  
\[ \alpha_n = \langle q_n | \mathbf{H} | q_n \rangle, \beta_n = \langle q_{n-1} | \mathbf{H} | q_n \rangle, \]  

(16)
The vectors formed by Eqs. (16) form the reduced Lanczos basis set, and the tridiagonal matrix with elements $\langle q_m | H | q_n \rangle$ takes the place of the matrix $H$ that appears in Eqs. (12). For $N$th order, in the sense of Eqs. (15), the reduced dimensional form of $H$ is

$$H_N = \begin{bmatrix}
\alpha_0 & \beta_0 & 0 & 0 \\
\beta_0 & \alpha_1 & \beta_1 & 0 \\
0 & \beta_1 & \alpha_2 & 0 \\
\vdots & \vdots & \vdots & \ddots \\
0 & 0 & 0 & \alpha_{N-1} \\
0 & 0 & 0 & \beta_{N-1} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix} \text{ (17)}$$

Making use of the reduced Lanczos representation, we can apply Eqs. (12a)–(12b), if we replace $z$ by $\Delta z$. The solution for the field over a single propagation step is then

$$\psi(\Delta z) = U^{-1} \exp[ih(1 - (1 + 2\beta_0/k)^{1/2})\Delta z]U\psi(0), \text{ (18a)}$$

$$\psi'(\Delta z) = U^{-1} \exp(-i\beta_0/\Delta z)U\psi'(0), \text{ (18b)}$$

where the matrices $\beta_0'$ and $U$ are determined by diagonalization of matrix (17). Note that if exactly $N$ modes are present in the field, the Lanczos propagation will be exact for order $N - 1$, and the $\Delta z$ in Eqs. (18) can be replaced by $z$.

### 6. SELECTION OF A PROPAGATION STEP

The selection of the propagation step $\Delta z$ in conjunction with the order $N$ of the Lanczos procedure will determine the accuracy of the solution. One diagnostic to monitor, which indicates the solution accuracy in a convergence sense, is the projection of $\psi(\Delta z)$ on $q_0$, the highest-order Lanczos vector. If this projection is smaller than some small preset value, one can conclude that the propagation step and the Lanczos order are appropriately matched. Thus one can control the solution error by adaptively altering $\Delta z$ to keep the projection of $\psi(\Delta z)$ on $q_0$ small.

Another constraint on $\Delta z$ that can be used to select the propagation step adaptively is that the step should be sufficiently small to allow all the phases in either Eq. (18a) or (18b) to be followed accurately. These conditions can be written as

$$\Delta z < \frac{\pi}{4} \min \left| \frac{1}{\Re[h(1 + 2\beta_0/k)^{1/2}]} \right|, \text{ (19a)}$$

$$\Delta z < \frac{\pi}{4} \min \left| \frac{1}{\beta_0'} \right|, \text{ (19b)}$$

which allow for sampling the complex exponentials in Eqs. (18a) and (18b), respectively, eight times per cycle. In our studies we use conditions (19) as default conditions, but we have found that accurately modeling the early stages of propagation may require even smaller steps following the launching of a beam into a rib waveguide.

When

$$2\beta_0'k < -1, \text{ (20)}$$

$(1 + 2\beta_0'/k)^{1/2}$ becomes pure imaginary, and the corresponding eigenvector attenuates exponentially. This can be understood as follows. If we approximate a radiation mode by a plane wave and take $n = n_o$, we have from Eq. (7)

$$2\beta_0'k = -(\kappa_x^2 + \kappa_y^2)/k^2, \text{ (21)}$$

where $\kappa_x$ and $\kappa_y$ are the applicable transverse spatial frequencies. Since the angle between the direction of propagation of the corresponding plane wave and the $z$ axis is

$$\theta = \sin^{-1}(\kappa_x^2 + \kappa_y^2)^{1/2}/k, \text{ (22)}$$

condition (20) implies that the radiation mode in question cannot propagate in the forward direction and therefore must attenuate exponentially.

During the transient phases of propagation in rib waveguides the sharp transitions in the refractive index generate high spatial frequency components in the field. The highest frequencies generated will be the highest frequencies that can be supported on the grid, that is, $K_{\text{max}} = \sqrt{2\pi/\Delta x}$. Through this filtering of the nonpropagating spatial frequencies, the iterative Lanczos reduction method applied to the Helmholtz equation ensures that the field will be band limited and, as a consequence, that its variation will be smooth. The paraxial wave equation, on the other hand, imposes no such restriction on which modes can propagate and thus can lead to short propagation steps, possible aliasing errors, and inaccurate transient fields.

### 7. SOLUTION FOR A BOUND MODE BY ITERATION

By successive applications of the Lanczos diagonalization procedure, in a sequence of steps suggestive of propagation, one can determine to high accuracy the bound mode and its eigenvalue for a monomode waveguide. This procedure is in the spirit of the traditional use for which the Lanczos method was intended, namely, the determination of the dominant eigenvalues of large matrices.

The iteration procedure works as follows. Let $\psi(0)$ be the starting field, and let the function values of this field on the grid represent the components of the vector $q_0$. One constructs $q_1, q_2, \ldots, q_N$ and the matrix $H_N$ from Eq. (16). Diagonalization of $H_N$ generates eigenvalues $\beta_0', \beta_1', \ldots, \beta_N'$.

![Fig. 1. Schematic diagram of rib waveguide geometry.](image)

<table>
<thead>
<tr>
<th>Table 1. Rib Waveguide Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>
where the ordering is such that

$$\beta_0' > \beta_2 > \ldots > \beta_N'$$.

For a monomode waveguide $\beta_0'$ will correspond to the bound mode and will be positive, while the remaining eigenvalues will be negative. Diagonalization also generates the transformation matrix $U$, which can be expressed in terms of the eigenvectors of $H_N$ by

$$U = [u_0 u_1 \ldots u_N]. \tag{23}$$

Assuming that $\beta_0'$ and $u_0$ are good approximations to the eigenvalue and eigenvector for the bound mode, one can discard the other eigenvectors and let $u_0$ serve as the input field for the next iteration. It is necessary, however, to express $u_0$ in terms of the $q$ vectors. Thus the iteration scheme can be expressed in the form

$$q^{n+1} = \sum_{n=0}^{N} [u_0]_n q_n^m, \tag{24}$$

where $m$ is an iteration number and $[u_0]_n$ represents a component of $u_0$ in the $N+1$ dimensional subspace in which $H_N$ is defined. The process leads to a successively weaker presence of all modes other than the desired bound mode. The procedure can be repeated until the solution has converged to the desired degree of accuracy, defined by the condition

$$\left| \frac{\beta_0^{m+1} - \beta_0^m}{\beta_0^m} \right| < \epsilon. \tag{25}$$
For the examples reported here we were able to achieve machine accuracy in iteration numbers varying from a few tens to a few hundreds and involving costs lower than would be incurred for a moderate-length propagation run.

For a directional coupler formed from two monomode waveguides there are two bound modes: a symmetric mode and an antisymmetric mode. Either mode can be computed by the above procedure by taking as the starting field either a symmetric or an antisymmetric function.

### 8. COMPUTATIONAL RESULTS

#### Waveguide Structures and Parameters

The basic rib waveguide structure is displayed in Fig. 1, and the relevant parameters are given in Table 1. The parameters were first used in the detailed study described in Ref. 5. A free-space wavelength of $\lambda = 1.55\, \mu\text{m}$ was used, and the reference index $n_0$ was set equal to $n_2$. In all cases the grid was set up so that refractive-index boundaries would lie midway between grid points. Grid boundaries were positioned so that the field would be small at the grid boundaries in both direct and wave number spaces. For some cases an absorbing mask was applied to the field near the boundaries in direct space, after a Lanczos propagation step was completed. This was done to avoid incorporating the absorber into the operator $\mathbf{H}$, which our development requires to be Hermitian.

#### Dynamic Evolution of Field Distributions as Determined by the Helmholtz Equation

Figure 2 shows the evolution of isointensity contour patterns at 1-\mu\text{m} intervals after a 0.3-\mu\text{m}, 1/e-intensity-

---

Fig. 3. Comparison between Helmholtz (left-hand column) and paraxial (right-hand column) beam propagation for evolution of a beam following injection of overfilling Gaussian beam into structure 1 rib waveguide.
radius Gaussian beam is launched into a structure 1 rib waveguide. The calculation was performed on a 128 × 128 grid with Δx = Δy = 0.1 μm. Initially the beam expands by diffraction until it collides with the dielectric-air interfaces. Subsequent reflections from the boundaries of the rib create standing wave patterns and cause energy to propagate downward into the substrate.

Figure 3 compares the propagations of intensity patterns determined by the Helmholtz and paraxial wave equations, following the launching of a Gaussian beam with a 1/e intensity radius of 1.0 μm into a structure 1 waveguide. The grid parameters are the same as for Fig. 2. Patterns are separated by 1 μm of propagation distance, and the isointensity contours are logarithmic with 20 contours spanning 10 orders of magnitude. Although the contour patterns for both types of propagation can be expected after a sufficient propagation distance to settle into the same bound-mode distribution, the patterns differ markedly in the early stages when radiation modes dominate the propagation.

Figure 4 demonstrates the transfer of energy between waveguides in a coupler formed from two waveguides of structure 2 with centers separated by 2 μm. Calculations were performed on a 128 × 64 grid with Δx = 0.15 μm and Δy = 0.05 μm. The field launched into the left-hand waveguide was the field corresponding to the single bound mode as determined by the iteration procedure in Section 7. Isointensity contours are displayed for propagation distance increments of one quarter of the coupling length Lc, which is 785 μm for this grid.

![Fig. 4. Energy transfer between waveguides formed from structure 2 waveguides. Mode of single waveguide was launched into left arm at z = 0. Distances are measured in coupling lengths Lc.](image-url)
Results for Bound Modes of Individual Waveguides and Couplers

The bound-mode propagation constants can be determined by propagating the beam a sufficient distance, calculating a field correlation function, and applying the spectrum analysis technique of Ref. 12. It was found, however, that when only one mode is at issue, the iteration procedure described in Section 7 is more accurate, is more efficient, and gives the mode eigenfunction as well. For each of the single-rib structures listed in Table 1 the bound Helmholtz eigenvalue was calculated by using the iteration method of Section 7. No attempt was made to

Fig. 5. Isointensity contours for bound mode of structure 2 waveguide, determined by iteration technique.

![Isointensity contours for bound mode of structure 2 waveguide](image)

Fig. 6. Isointensity contours for bound mode of structure 3 waveguide, determined by iteration technique.

![Isointensity contours for bound mode of structure 3 waveguide](image)
optimize the calculations with regard to computation time, since the objective of the calculations was accuracy. Computations were arbitrarily performed to ninth order, that is, with 10 Lanczos vectors. But accurate results could also have been obtained for lower orders with more iteration steps. For the tightly bound mode corresponding to structure 1 the eigenvalue converged to machine accuracy in fewer than 30 steps. For structures 2 and 3 the iteration steps have converged with respect to grid size and parameters, as, for example, in the case in Table 2 for the single structure 2 waveguide, the accuracy of the final answer is firmly established.

There is no objective way to compare the accuracies of the different methods, since the correct propagation constants are not known. But the Lanczos method has an advantage over the others for assessing accuracy; namely, for a given sampling grid the accuracy of the propagation constants can be specified in advance by employing relation (25). This means that when the resulting solution has converged with respect to grid size and parameters, as, for example, in the case in Table 2 for the single structure 2 waveguide, the accuracy of the final answer is firmly established.

9. SUMMARY AND CONCLUSION
We have reviewed the iterative Lanczos reduction method for solving both the Helmholtz and paraxial wave equations, and we have described an iteration scheme for accurately determining the mode and propagation constant of a monomode waveguide.

Calculational results for the early stages of propagation following the injection of a beam into a semiconductor rib waveguide indicate substantial differences between Helmholtz and paraxial wave propagation, just as might be expected. We have computed propagation constants and mode fields for three canonical waveguide structures, as well as coupling lengths for the corresponding couplers using the iteration scheme. The iteration scheme is capable in principle of producing propagation constants to machine accuracy for a specified grid, which we point out is a significant advantage of the method.

ACKNOWLEDGMENT
This study was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48.

REFERENCES AND NOTES
2. T. J. Park and J. C. Light, “Unitary quantum time evolution...


