Variational Finite-Element Solution for Dissipative Waveguides and Transportation Application

ALASTAIR D. McAULAY, MEMBER, IEEE

Abstract—**A procedure is developed for determining the complex propagation constants and associated complex electromagnetic fields as a function of frequency for electromagnetic waves propagating along an inhomogeneous waveguide composed of dissipative materials and having a complicated shape. The wave equation, which is complex because of the presence of dissipative materials, is transformed for computer solution into a matrix eigenvalue equation by the application of the Rayleigh–Ritz variational method in conjunction with the finite-element method. The results are reviewed for several simple dissipative waveguides for which analytical results are computed for comparison. A novel proposal is then investigated in which a railroad track acts as a surface waveguide for a rapid-transit collision-avoidance system. The results illustrate the usefulness of the numerical method developed and suggest that the modified steering rail warrants further investigation for rapid-transit systems.**

I. Introduction

THE PROPAGATION of electromagnetic waves is investigated along waveguides composed of dissipative materials and having a complicated shape. In particular, the complex propagation constants and their associated complex electric and magnetic fields are computed as a function of frequency for the lower order modes. Previously, such dissipative waveguides were analyzable analytically only in the case of low-order modes, simple shapes, and involving materials having either very low dissipation, such as a dissipative dielectric rod [1], or very high conductivity, such as a conducting wire [2].

Earlier numerical methods [3]–[10] enabled the analysis of waveguides having complicated shapes and containing inhomogeneous materials, but were restricted to lossless materials. Many waveguides possess low levels of attenuation and the effect of dissipation may be calculated accurately from the lossless fields by perturbation methods [1]. However, the lossless waveguide techniques are inadequate where loss is significant, or where a surface wave propagates only because of the presence of dissipation, such as in a conducting wire [2] or Zenneck's surface wave [11].

The effect of dissipative materials can be reduced to conversion of the electromagnetic wave equation into a complex wave equation. The Rayleigh–Ritz variational method in conjunction with the finite-element method is applied for the first time to the complex wave equation. The object of the analysis is to convert these equations into a matrix eigenvalue equation suitable for computer solution. The procedure developed differs from the lossless case: first, it is necessary to find a real functional for minimization, and, second, the complex matrix eigenvalue equation must be solved for real eigenvalues and complex eigenvectors. Assuming a specified phase velocity, the complex propagation constants and their associated complex fields may then be determined for the possible modes of propagation along a waveguide. Solving the problem for different phase velocities enables the solutions to be obtained as a function of frequency. The author has obtained an identical result [13], [14] by applying the method of weighted residuals [12] in place of the Rayleigh–Ritz method.

The numerical method was verified by applying it to waveguides for which some analytical solutions were available for comparison [13], [14]. A brief summary of these results is presented. The method was then applied to investigate a novel technique, proposed by the author, for use in headway control in track-guided transportation systems [15]. The problem of controlling the spacing of rapid-transit vehicles in a safe manner is one of considerable interest to the Department of Transportation [16].

II. Derivation of Equations

The wave is assumed to propagate in the positive z-direction in a time-harmonic manner, and there is no other z-dependence because a waveguide is being considered. Then

$$\mathscr{E} = \mathrm{Re}\,[\boldsymbol{E}e^{j(\gamma z-\omega t)}] \quad \mathscr{H} = \mathrm{Re}\,[\boldsymbol{H}e^{j(\gamma z-\omega t)}] \qquad (1)$$

where the propagation constant $\gamma = \beta + j\alpha$, the angular frequency ω is real and the components of the vectors $\boldsymbol{E}$ and $\boldsymbol{H}$ are functions of the coordinates transverse to the direction of propagation along the guide. Substituting (1) into the source-free wave equations produces the z-component transverse-harmonic wave equations

$$(\nabla^2 + k^2 - \gamma^2)E_z = 0 \quad (\nabla^2 + k^2 - \gamma^2)H_z = 0 \qquad (2)$$

where ∇^2 is now the two-dimensional Laplacian operator and the intrinsic propagation constant is

$$k = \omega\sqrt{\mu\varepsilon}. \qquad (3)$$

Maxwell's equations may be used to obtain the tangential components of $\boldsymbol{E}$ and $\boldsymbol{H}$ from the z components; therefore, only the equations involving the z components need to be solved.

Manuscript received April 22, 1976; revised August 9, 1976. Part of this work was performed while the author was with Carnegie-Mellon University, and was funded by the Department of Transportation.

The author is with Honeywell Marine Systems, 5303 Shilshole Ave. N.W., Seattle, WA 98107.

The presence of conductivity σ may be included as an additional imaginary part of the material permittivity

$$\varepsilon = \varepsilon' + j\varepsilon'', \qquad \varepsilon'' = \sigma/\omega. \tag{4}$$

The intrinsic propagation constant k in (3) is now complex and, consequently, in general γ, E_z, and H_z of (2) are also complex.

The transverse-harmonic wave equation and Maxwell's equations may be written in nondimensional form [13] by defining a fixed length a which is normally a suitable dimension in the problem. ∇, k, and γ are multiplied by a to produce nondimensional quantities; μ and ε are replaced by the quantities relative to a vacuum μ_r and ε_r; E_z and H_z are similarly divided by arbitrary reference fields; and ω is multiplied by a/c to obtain the nondimensional ω. The variables are assumed to be nondimensional for the remainder of this paper unless otherwise specified.

The finite-element method and its applications are well documented in structural engineering [17], elasticity [18], fluid mechanics, heat, and mass transfer [12]. The subdomain principle is utilized in which the domain of the equation to be solved is divided into separate regions or subdomains. In the case of the transverse electromagnetic wave equation, the domain is the area of cross-section of the waveguide. Triangular subdomains are chosen for convenience. The unknown solution functions in (2) are now approximated in each subdomain by the sum of a set of polynomials [19], and a basis transformation [13] is applied to produce

$$E_z = \sum_{i=1}^{n} a_i\phi_i(x,y) \quad H_z = \sum_{i=1}^{n} b_i\phi_i(x,y) \tag{5}$$

where the $\phi_i(x,y)$ are selected as linear functions of x and y, having a value in only one triangle and unity value at only one node of the triangle, a_i are the unknown values of the electric field E_z at the n triangle nodes in the cross section, and b_i are the unknown values of the magnetic field H_z at these nodes.

The a_i and b_i are to be determined so that (5) is in some way the best approximation to the solution of (2). For a first-order polynomial approximation, each of the k triangles has three nodes; therefore, $n = 3k$. A first-order polynomial is selected in order to fit more easily into complicated boundary shapes. Other methods are available for boundaries of known curvature [20].

The Rayleigh–Ritz variational method requires a real functional for the purpose of extremization. One method of forming a suitable functional is to split the complex wave equations (2) into real and imaginary parts. Using (3), defining $m = \mu_r\varepsilon_r - (\gamma/\omega)^2$, and using single- and double-prime superscripts for the real and imaginary parts of the variables, respectively, produces

$$\nabla^2\boldsymbol{I}\boldsymbol{\Phi} + \omega^2\boldsymbol{M}\boldsymbol{\Phi} = 0 \tag{6}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\boldsymbol{\Phi} = \begin{bmatrix} E_z' \\ E_z'' \\ H_z' \\ H_z'' \end{bmatrix} \quad \boldsymbol{M} = \begin{bmatrix} m' & -m'' & 0 & 0 \\ m'' & m' & 0 & 0 \\ 0 & 0 & m' & -m'' \\ 0 & 0 & m'' & m' \end{bmatrix}. \tag{7}$$

Equation (6) has the form

$$\boldsymbol{L}\boldsymbol{\Phi} = 0 \tag{8}$$

where the linear differential operator

$$\boldsymbol{L} = \nabla^2\boldsymbol{I} + \omega^2\boldsymbol{M} \tag{9}$$

acts on the function $\boldsymbol{\Phi}(x,y)$ and is defined in the surface region Ω. The boundary conditions, Dirichlet, Neumann, or mixed may be written similarly

$$\boldsymbol{l}_i\boldsymbol{\Phi}|_{r_i} = \boldsymbol{\alpha}_i(x,y) \tag{10}$$

where $\boldsymbol{\alpha}_i$ is the known function along the ith segment r_i of the boundary r, and $\boldsymbol{l}_i$ is a linear operator corresponding to the ith segment.

In the case of lossless waveguides, the corresponding operator $\boldsymbol{L}$ is self-adjoint, and it can be proved that a variation formula is

$$F = \langle\boldsymbol{\Phi},\boldsymbol{L}\boldsymbol{\Phi}\rangle. \tag{11}$$

However, in this case because losses are present, the operator of (9) is not self-adjoint. This is shown by substituting (7) into (9) and establishing that

$$\langle\boldsymbol{u},\boldsymbol{L}\boldsymbol{v}\rangle \neq \langle\boldsymbol{v},\boldsymbol{L}\boldsymbol{u}\rangle \tag{12}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are arbitrary vectors of suitable dimension. The variational formula for a non self-adjoint operator is

$$F = \langle\boldsymbol{\Phi}^a,\boldsymbol{L}\boldsymbol{\Phi}\rangle \tag{13}$$

where $\boldsymbol{\Phi}^a$ is the adjoint variable of $\boldsymbol{\Phi}$ [12]. The integral over the area is used as the inner product for the case of the transverse wave equations. Equation (13) is shown to be a variational formula for the operator of (9) by considering a perturbation

$$F + \delta F = \int_S (\boldsymbol{\Phi}^a + \delta\boldsymbol{\Phi}^a)\boldsymbol{L}(\boldsymbol{\Phi} + \delta\boldsymbol{\Phi})\, dS. \tag{14}$$

Equation (13) is subtracted, second-order terms are neglected, and the Laplacian is expanded using integration by parts twice

$$\delta F = \int_S \delta\boldsymbol{\Phi}^a\boldsymbol{L}\boldsymbol{\Phi}\, dS + \int_S \delta\boldsymbol{\Phi}\,\nabla^2\boldsymbol{I}\boldsymbol{\Phi}^a\, dS + \oint_\tau (\boldsymbol{\Phi}^a\nabla\boldsymbol{I}\delta\boldsymbol{\Phi} - \delta\boldsymbol{\Phi}\nabla\boldsymbol{I}\boldsymbol{\Phi}^a)\, d\tau + \omega^2\int_S \delta\boldsymbol{\Phi}\boldsymbol{M}^T\boldsymbol{\Phi}^a\, dS. \tag{15}$$

This may be written

$$\delta F = \langle\delta\boldsymbol{\Phi}^a\boldsymbol{L}\boldsymbol{\Phi}\rangle + \langle\delta\boldsymbol{\Phi}\boldsymbol{L}^a\boldsymbol{\Phi}^a\rangle + B(\boldsymbol{\Phi}^a,\delta\boldsymbol{\Phi}) \tag{16}$$

where the last term contains the contour integral boundary terms and where the identity was used ($\boldsymbol{L}$ adjoint) = ($\boldsymbol{L}$ transpose) or $\boldsymbol{L}^a = \boldsymbol{L}^T$ for real L.

The functional equation (13) is variational if and only if the variation resulting from a perturbation of the variables

$$\delta F = 0. \tag{17}$$

Consequently, from (16) it can be seen that (17) is satisfied if $\boldsymbol{L}\Phi = 0$, which is the original (8), and

$$\boldsymbol{L}^a\Phi^a = 0 \tag{18}$$

which is the adjoint of this equation and in addition the boundary term

$$B(\Phi^a,\delta\Phi) = 0. \tag{19}$$

Equation (19) requires that Dirichlet, Neumann, or mixed boundary conditions are satisfied. The variational formula in the case of the wave equations (6) is now written using (13)

$$F = \int_S \Phi^a \nabla^2 \boldsymbol{I}\Phi \, dS + \int_S \omega^2 \Phi^a \boldsymbol{M}\Phi \, dS. \tag{20}$$

Equation (20) is expanded using Green's theorem because the Laplacian of a linear function (5) vanishes, which is not a realistic approximation to the Laplacian of the actual field

$$F = -\int_S \nabla\Phi^a \cdot \nabla\Phi \, dS + \oint_\tau \Phi^a \cdot \frac{\partial\Phi}{\partial n} \, d\tau + \int_S \omega^2 \Phi^a \boldsymbol{M}\Phi \, dS. \tag{21}$$

Maxwell's equations, in normalized form and split into real and imaginary parts, are used to express the derivatives which are discontinuous at the interface between different media, $\partial\Phi/\partial n$, in terms of continuous derivatives $\partial\Phi/\partial\tau$

$$\frac{\partial\Phi}{\partial n} = \boldsymbol{P}\frac{\partial\Phi}{\partial\tau} \tag{22}$$

where

$$\frac{\partial\Phi}{\partial\tau} = \begin{bmatrix} \partial E_z'/\partial\tau \\ \partial E_z''/\partial\tau \\ \partial H_z'/\partial\tau \\ \partial H_z''/\partial\tau \end{bmatrix} \quad \boldsymbol{P} = \begin{bmatrix} 0 & 0 & -Q' & Q'' \\ 0 & 0 & -Q'' & -Q' \\ R' & -R'' & 0 & 0 \\ R'' & R' & 0 & 0 \end{bmatrix} \tag{23}$$

and

$$Q' = \frac{(\gamma/\omega)'\varepsilon_r' + (\gamma/\omega)''\varepsilon_r''}{|\varepsilon_r|^2} \tag{24a}$$

$$Q'' = \frac{(\gamma/\omega)''\varepsilon_r' - (\gamma/\omega)'\varepsilon_r''}{|\varepsilon_r|^2}$$

$$R' = \frac{(\gamma/\omega)'\mu_r' + (\gamma/\omega)''\mu_r''}{|\mu_r|^2} \tag{24b}$$

$$R'' = \frac{(\gamma/\omega)''\mu_r' - (\gamma/\omega)'\mu_r''}{|\mu_r|^2}.$$

Hence, (21) may be written

$$F = -\int_S \nabla\Phi^a \cdot \nabla\Phi \, dS + \oint_\tau \Phi^a \boldsymbol{P}\frac{\partial\Phi}{\partial\tau} \, d\tau + \int_S \omega^2 \Phi^a \boldsymbol{M}\Phi \, dS. \tag{25}$$

The Rayleigh–Ritz procedure is followed. Approximation functions are substituted for the functions Φ and Φ^a in the functional equation (25), which is then extremized by differentiation with respect to the unknown coefficients. The approximation functions are obtained by splitting (5) into real and imaginary parts, and further approximation functions are defined for the adjoint variables; for example,

$$E_z'^a = \sum_{i=1}^{n} a_i'^a \phi_i(x,y). \tag{26}$$

The summation limits are assumed to be $i = 1$ to n for the remainder of this paper, unless otherwise stated.

The first term of (25) becomes

$$F_1 = -\int_S [(\Sigma a_i'^a \nabla\phi_i)(\Sigma a_i' \nabla\phi_i) + (\Sigma a_i''^a \nabla\phi_i)(\Sigma a_i'' \nabla\phi_i) + (\Sigma b_i'^a \nabla\phi_i)(\Sigma b_i' \nabla\phi_i) + (\Sigma b_i''^a \nabla\phi_i)(\Sigma b_i'' \nabla\phi_i)] \, dS. \tag{27}$$

Differentiation with respect to the coefficients a_i', a_i'', b_i', b_i'', $a_i'^a, a_i''^a, b_i'^a, b_i''^a$, and defining a matrix

$$s_{ji} = \int_S \nabla\phi_j \cdot \nabla\phi_i \, dS \tag{28}$$

which may be precomputed for a given triangle shape [21], produces

$$\begin{bmatrix} \partial F_1/\partial \boldsymbol{x}^a \\ \partial F_1/\partial \boldsymbol{x} \end{bmatrix} = -\begin{bmatrix} s_{ji}\boldsymbol{I} & 0 \\ 0 & (s_{ji}\boldsymbol{I})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^a \end{bmatrix} \tag{29}$$

where

$$\boldsymbol{x} = [a_i', a_i'', b_i', b_i'']^T \qquad \boldsymbol{x}^a = [a_i'^a, a_i''^a, b_i'^a, b_i''^a]^T. \tag{30}$$

In a similar manner, the second term of (25) becomes

$$\begin{bmatrix} \partial F_2/\partial \boldsymbol{x}^a \\ \partial F_2/\partial \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}w_{ji} & 0 \\ 0 & (\boldsymbol{P}w_{ji})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^a \end{bmatrix} \tag{31}$$

after differentiating with respect to the coefficients and defining a matrix

$$w_{ji} = \oint_\tau \phi_j \frac{\partial\phi_i}{\partial\tau} \, d\tau \tag{32}$$

which may be precomputed [22].

Similarly, the third term of (25), using (7), becomes

$$\begin{bmatrix} \partial F_3/\partial \boldsymbol{x}^a \\ \partial F_3/\partial \boldsymbol{x} \end{bmatrix} = \omega^2 \begin{bmatrix} \boldsymbol{M}t_{ij} & 0 \\ 0 & (\boldsymbol{M}t_{ij})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^a \end{bmatrix} \tag{33}$$

after differentiating with respect to the coefficients and defining another precomputable matrix [21]

$$t_{ji} = \int_S \phi_j \phi_i \, dS. \tag{34}$$

The combination of the three differentiated terms (29), (31), and (33), corresponding to the three terms in (25), are set to zero in order to extremize the variational functional F

$$-\begin{bmatrix} s_{ji}\boldsymbol{I} & 0 \\ 0 & (s_{ji}\boldsymbol{I})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^a \end{bmatrix} + \begin{bmatrix} \boldsymbol{P}w_{ji} & 0 \\ 0 & (\boldsymbol{P}w_{ji})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^a \end{bmatrix} = -\omega^2 \begin{bmatrix} \boldsymbol{M}t_{ji} & 0 \\ 0 & (\boldsymbol{M}t_{ji})^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^a \end{bmatrix}. \tag{35}$$

The lower matrix equation of (35) is the adjoint of the upper equation, and consequently has the same eigenvalues and adjoint eigenvectors. Thus only one equation need be solved

$$[s_{ji}\boldsymbol{I} - \boldsymbol{P}w_{ji}]\boldsymbol{x} = \omega^2 \boldsymbol{M}t_{ji}\boldsymbol{x}. \tag{36}$$

A matrix eigenvalue equation for a cross section containing k triangles may be constructed using (36), because the approximation function (5) for each triangle is zero in all other triangles. $\boldsymbol{S},\boldsymbol{W},\boldsymbol{T},\boldsymbol{a}',\boldsymbol{a}'',\boldsymbol{b}',\boldsymbol{b}''$ are used for the matrices where previously only the matrix elements were used and the subscripts now refer to the matrix or vector triangle number.

$$\begin{bmatrix} \boldsymbol{S}_1 \searrow \boldsymbol{S}_k & \boldsymbol{0} & -\boldsymbol{Q}_1'\boldsymbol{W} \searrow -\boldsymbol{Q}_k'\boldsymbol{W} & \boldsymbol{Q}_1''\boldsymbol{W} \searrow \boldsymbol{Q}_k''\boldsymbol{W} \\ \boldsymbol{0} & \boldsymbol{S}_1 \searrow \boldsymbol{S}_k & -\boldsymbol{Q}_1''\boldsymbol{W} \searrow -\boldsymbol{Q}_k''\boldsymbol{W} & -\boldsymbol{Q}_1'\boldsymbol{W} \searrow -\boldsymbol{Q}_k'\boldsymbol{W} \\ \boldsymbol{R}_1'\boldsymbol{W} \searrow \boldsymbol{R}_k'\boldsymbol{W} & -\boldsymbol{R}_1''\boldsymbol{W} \searrow -\boldsymbol{R}_k''\boldsymbol{W} & \boldsymbol{S}_1 \searrow \boldsymbol{S}_k & \boldsymbol{0} \\ \boldsymbol{R}_1''\boldsymbol{W} \searrow \boldsymbol{R}_k''\boldsymbol{W} & \boldsymbol{R}_1'\boldsymbol{W} \searrow \boldsymbol{R}'_k\boldsymbol{W} & \boldsymbol{0} & \boldsymbol{S}_1 \searrow \boldsymbol{S}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_k' \\ \boldsymbol{a}_1'' \\ \vdots \\ \boldsymbol{a}_k'' \\ \boldsymbol{b}_1' \\ \vdots \\ \boldsymbol{b}_k' \\ \boldsymbol{b}_1'' \\ \vdots \\ \boldsymbol{b}_k'' \end{bmatrix}$$

$$= \omega^2 \begin{bmatrix} \boldsymbol{m}_1'\boldsymbol{T}_1 \searrow \boldsymbol{m}_k'\boldsymbol{T}_k & -\boldsymbol{m}_1''\boldsymbol{T}_1 \searrow -\boldsymbol{m}_k''\boldsymbol{T}_k & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{m}_1''\boldsymbol{T}_1 \searrow \boldsymbol{m}_k''\boldsymbol{T}_k & \boldsymbol{m}_1'\boldsymbol{T}_1 \searrow \boldsymbol{m}_k'\boldsymbol{T}_k & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{m}_1'\boldsymbol{T}_1 \searrow \boldsymbol{m}_k'\boldsymbol{T}_k & -\boldsymbol{m}_1''\boldsymbol{T}_1 \searrow -\boldsymbol{m}_k''\boldsymbol{T}_k \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{m}_1''\boldsymbol{T}_1 \searrow \boldsymbol{m}_k''\boldsymbol{T}_k & \boldsymbol{m}_1'\boldsymbol{T}_1 \searrow \boldsymbol{m}_k'\boldsymbol{T}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_k' \\ \boldsymbol{a}_1'' \\ \vdots \\ \boldsymbol{a}_k'' \\ \boldsymbol{b}_1' \\ \vdots \\ \boldsymbol{b}_k' \\ \boldsymbol{b}_1'' \\ \vdots \\ \boldsymbol{b}_k'' \end{bmatrix}. \tag{37}$$

The $\boldsymbol{a}$ and $\boldsymbol{b}$ vectors in (37) form a vector of length $3 \times 4 \times k$ representing the unknown complex electric- and magnetic-field values at each vertex of the k triangles. When the triangles are brought together, many of these unknowns refer to the same points and must therefore be equal. This continuity condition is accomplished by initially assigning numbers to each node of the cross section rather than each vertex of each triangle. The submatrices in (37) are then entered into the correct positions of the final matrix directly.

At external-boundary node points, where $E_z = 0$ or there are lines of symmetry for the tangential electric field, the appropriate electric-field values of a' and a'' are set to zero. Similarly, where $H_z = 0$, or there are lines of symmetry for the tangential magnetic field, the appropriate magnetic-field values of b' and b'' are set to zero. Consequently, some rows and their corresponding columns are omitted.

In the cases that the resulting eigenvalues ω^2 are real, they provide the frequencies of possible modes of propagation and the eigenvectors determine the complex electric and magnetic fields at the triangle nodes in the cross section.

The matrices involving $Q'W$ couple the real parts of the electric and magnetic field and are zero for TM and TE modes, and nonzero only for hybrid modes. The matrices involving double-prime quantities are zero when there are no dissipative materials present. The matrices involving $Q''W$ are nonzero only for hybrid modes in the presence of dissipative materials.

The result obtained is shown in [13] to be identical to that obtained by means of the method of weighted residuals [12], [14].

III. Numerical Procedure

The waveguide cross section is divided into triangular regions at whose vertices the complex electric and magnetic fields are determined. Consequently, a finer net is selected where one of these four functions is expected to vary rapidly. If too few triangles are used, or if they are positioned

badly, or truncation of the surface wave decaying to infinity is too great, the eigenvalues will be less accurate.

A program for the solution of the general matrix eigenvalue equation $\boldsymbol{Ax} = \lambda \boldsymbol{Bx}$ called the QZ algorithm [23] was used to compute the complex eigenvalues λ and complex eigenvectors x. The frequencies were determined from $\omega^2 = \lambda$, when λ was real.

The smallest positive eigenvalue for an arbitrary selection of α/ω and β/ω is generally complex because the matrix is unsymmetrical. However, the frequency was assumed to be real, and it is therefore necessary to iterate on α/ω or β/ω until the imaginary part of ω becomes negligible [13]. Similarly, the imaginary part of the dielectric, which depends on ω, will need to be iterated until $\varepsilon_r'' = \sigma_r/\omega$ for a medium of relative conductivity σ_r.

The numerical process involves discretization of a continuous field. Consequently, spurious modes are generated which are not realizable physical modes. In addition, some modes are lost, such as the higher order modes, whose field variations can no longer be adequately represented by a limited number of triangles. Neither of these features present a problem because the lowest order modes are the ones of interest and it was found that these are easily distinguished from the spurious modes which exhibit obvious nonphysical characteristics [13].

The variation of the electromagnetic field with time is provided by the fact that the electric and magnetic field are complex. For example, the electric field at a node, $E = a + bj$ produces

$$\begin{aligned} \mathscr{E}(t) &= \text{Re}\,[(a + bj)e^{j\omega t}] \\ &= a\cos\omega t - b\sin\omega t \\ &= \sqrt{a^2 + b^2}\cos(\omega t + \tan^{-1}(b/a)). \end{aligned} \tag{38}$$

Hence, at $\omega t = 0$, $\mathscr{E}(t) = a$; at $\omega t = \pi/2$, $\mathscr{E}(t) = -b$; and at $\omega t = -\tan^{-1}(b/a)$, $\mathscr{E}(t) = \sqrt{a^2 + b^2}$. The knowledge of the field variation with time enables damped transverse-wave motion to be detected.

IV. Verification Examples

In order to verify the accuracy of the previously mentioned numerical procedure, several relatively simple waveguide structures were investigated for which some analytical results may be calculated for comparison [1], [2], [24], [25]. These results are described in detail in [14]; consequently, only a brief review is presented here. As successively more complex guides were considered, more submatrices in (37) became nonzero.

The complex propagation constants and fields were determined as a function of frequency for a hollow circular waveguide filled with dissipative material in order to verify the iteration process for finding real frequencies. Typically, errors of less than 1 percent were obtained for specified propagation constants, a TM_{01} mode, and a 15 triangle approximation [14].

Hybrid modes may propagate along a dielectric rod and the fields decay to infinity transversely. The TM_{01}, TE_{01}, and HE_{11} mode propagation constants showed close agreement between the numerical and analytical computations [12] for a wide range of frequencies [14]. The numerical method was also shown to be capable of providing good results for a dielectric-coated conductor [14], [24].

A rod having both dielectric and conductive properties was investigated. Analytical methods enable solutions to be obtained for the lowest order modes and limited frequency ranges when the losses are small (perturbation methods) [1] or when the conductivity is high [2], [25]. In contrast, the numerical method developed here can handle all in-between cases. The numerical results show that, as the conductivity is increased, surface-wave propagation as a result of the dissipative surface phenomena increases while that due to internal reflection from side to side in the dielectric becomes less significant [14]. Both conducting rods in which the diameter is much less than the skin depth [2] and in which the diameter is much greater than the skin depth [25] were considered. In the latter case, surface-wave propagation is entirely due to the dissipative property on one side of the surface interface. In the former case, because of the internal reflection phenomena, the surface wave decays more rapidly in the transverse direction and differs more substantially from a plane wave. Good agreement was obtained between the numerical and analytical techniques even though the field decays to $1/e$ of the surface value in millionths of a meter inside the conductor and meters outside the conductor [14].

V. Application to Railroad Track

A. Approach

A method is proposed in which the track, used for steering a rapid-transit vehicle, is adapted for use as a surface waveguide [15]. Signals transmitted ahead of a vehicle are steered by the track and reflected from obstacles or vehicles ahead. In order to evaluate this proposal, the complex propagation constant for the lowest order mode and the associated electromagnetic fields are computed for a given track structure. The railroad rail was selected because it is the most common form of track guiding system and is still popular on new systems, such as the BART system in San Francisco and the subway in Washington, DC.

The surface wave is required to cling fairly closely to the rail in order not to strike objects alongside the track. Consequently, it was decided to modify the rail by the addition of a strip of dielectric material, as in Fig. 1. The surface wave is enhanced because the wave is partially trapped inside the dielectric by repeated internal reflection as the wave propagates. The lowest order mode which propagates is similar to the HE_{11} mode on a rod, the arrangement being similar to a dielectric image line.

A similar approach is possible for most track guiding systems because there is generally a steel guideway even when rubber-tired vehicles are used.

The complex propagation constant for the HE_{11} mode is computed to determine the phase distortion and attenuation, and the complex fields are determined for possible use

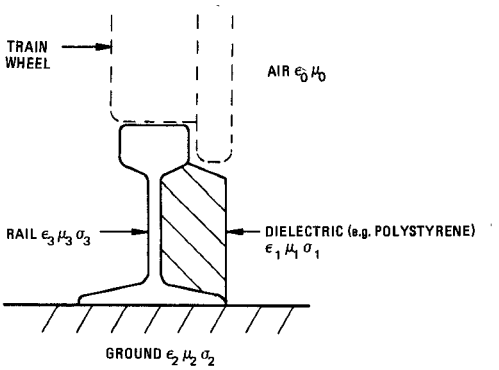

Fig. 1. Rail augmented with strip of Dielectric material resting on ground.

in antenna and reflector designs. The results, including those for a water saturated track, indicate that the method proposed warrants further investigation in the search for a more economic and efficient headway control and collision avoidance system.

B. Model

The surface waveguide characteristics of a rail modified by a dielectric strip are influenced by the properties of the four regions present: the rail, the dielectric strip, the air, and the ground.

The dielectric material chosen should have a high value of relative permittivity, a small loss tangent, and a low price. A commonly available material, polystyrene, was used for the analysis. There may be other more suitable materials which would produce better results. The polystyrene properties assumed were relative permittivity $\varepsilon_r = 2.56$ and loss tangent $\tan \delta = 0.00033$ in the range of frequencies considered [26].

The rail properties assumed were conductivity $\sigma = 0.6 \times 10^8$ mho/m, relative permittivity $\varepsilon_r = 1$, and relative permeability $\mu_r = 10$, the latter value being estimated from values at lower frequencies. It was expected that the loss in the rail would be small relative to the loss in the ground and the dielectric because of the high rail conductivity; therefore, initially this loss was neglected. A perturbation method was used later, and it was found that the loss due to the rail was less than 2 percent of that due to the ground and the dielectric together [13].

The properties of the materials directly beneath the rail are harder to determine. Normally, the rails rest on wooden sleepers and these rest on a ballast material made of gravel. The part of the wave below the rail is partially reflected by the sleepers.

The problem of periodic discontinuities is discussed by Hu [27] and is also relevant to the effect of gaps in nonwelded rails. The reflections caused by repetitive discontinuities is shown to have a compounded effect and is not just linearly additive. It was decided that a suitable model, expected to give the worst case results, would be to consider the rail resting directly on the ground. Electrical properties at 3×10^9 Hz for different ground materials are shown in Table I [26]: the sandy-dry and distilled-water cases were considered in order to illustrate the procedure and to observe the effect of a large variation in the values of the parameters.

TABLE I
GROUND MATERIAL PROPERTIES

	ϵ_r	TAN δ
SANDY DRY	2.5	0.0062
LOAMY DRY	2.44	0.0011
SNOW	1.2	0.00029
DISTILLED WATER	76.7	0.157

The air was considered to have a relative permittivity $\varepsilon_r = 1$, a relative permeability $\mu_r = 1$, and zero conductivity. The effect of fog and rain are not significant in the range of frequencies considered [28].

Symmetry of the field about a vertical plane along the center of the rail was assumed; consequently, only half the rail cross-section is considered. In practice, the dielectric might be placed on one side of the rail and the results would be slightly different. The manner in which the half cross-section is subdivided into triangles is shown in Fig. 2. The radial extent of the triangles was chosen so that a sufficient part of the field would be in the air and available for coupling with the vehicle and at the same time the field would not be extended too far into the lossy ground. A smaller and larger triangle spread were used for higher and lower frequencies, respectively. A triangle spread selected for a given frequency will result in increasingly truncated field representation as the frequency is lowered and increasingly poor triangle utilization as the frequency is raised.

C. Results

The numerical method was applied to the augmented rail model and the results obtained for the normalized phase constant versus normalized frequency are shown in Fig. 3. Examination of the HE_{11}-type field patterns for three layouts showed that a particular layout is only valid over a small range of frequencies. The true curve is expected to lie between the curves shown. The group velocity $d\omega/d\beta$, which determines the phase distortion of the signal, is seen to be approximately constant over the range of frequencies considered. Consequently, phase distortion is expected to be minimal at these frequencies.

The normalized attenuation constant divided by normalized frequency versus normalized frequency, Fig. 4, is extrapolated from the three dotted curves and agrees with the predicted behavior. At very high frequencies, the wave withdraws into the dielectric material and approaches a plane wave with a ratio of normalized propagation constant to normalized frequency of $\gamma/\omega = 1.6 + j0.000264$. At the lower frequencies, the wave spreads farther from the rail, which simplifies coupling with the vehicle, but causes an increase in attenuation because more of the field enters the ground.

The normalized attenuation constant was determined for a configuration in which the ground is considered lossless

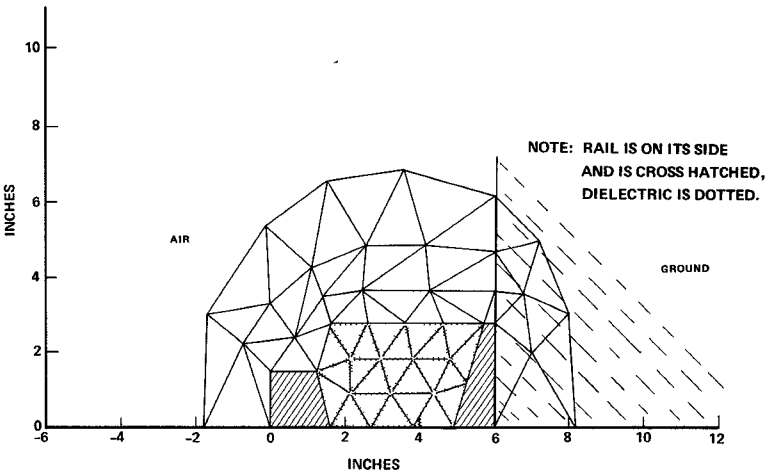

Fig. 2. Layout of triangles for application of finite-element method.

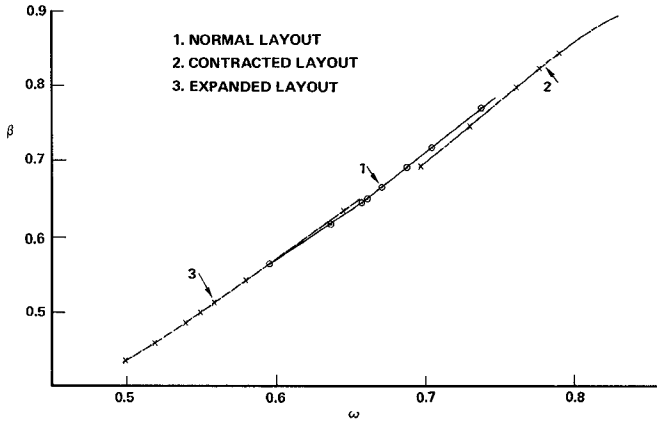

Fig. 3. Phase constant versus frequency for an augmented rail.

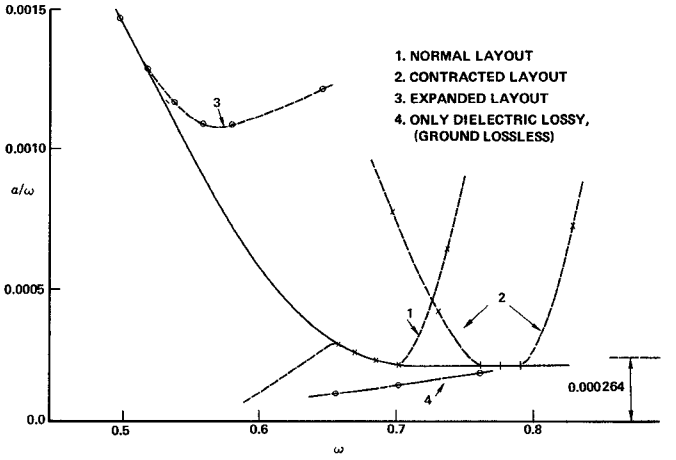

Fig. 4. Attenuation constant divided by frequency versus frequency for an augmented rail.

(curve 4 of Fig. 4) in order to determine the apportionment of the attenuation between that caused by dissipation in the dielectric and that caused by dissipation in the ground. The attenuation caused by dissipation in the dielectric decreases as the frequency is lowered. At normalized frequencies below $\omega = 0.68$, the loss caused by the ground becomes dominant, while above this frequency the dielectric loss becomes significant.

Figs. 5 and 6 show, superimposed on the cross section of a rail, the real part of the z component of the electric and magnetic fields at a frequency 1.246 GHz ($\omega = 0.659$) near the center of the range considered. The imaginary part of the electric and magnetic fields represent the electric and magnetic fields at a time $\omega t = \pi/2$ later than the real part and are similar in appearance, having a smaller magnitude because the attenuation involved is small. A ratio of normalized propagation constant to normalized frequency of $\gamma/\omega = \beta/\omega + j\alpha/\omega = 1.145 + j0.000306$ was used. At this ratio of propagation constant to frequency the imaginary part of the eigenvalue is small (0.000054) relative to the value when either α/ω or β/ω are changed only slightly. The field patterns are seen to be similar to those for an HE_{11} model on a lossy dielectric rod [14]. The field extends into the ground and the transverse distance from the surface of the dielectric to the point where the z component of the electric field has decayed to $1/e$ of the value at the surface, is about 0.67 in. The z component of the magnetic field in the direction of propagation has a maximum value along the beveled edge of the dielectric. The transverse magnetic and electric fields are obtainable from Maxwell's equations

$$h_t = -\frac{j\beta}{k_c^2}\nabla_t h_z - \frac{j\omega\varepsilon}{k_c^2}(\hat{a}_z \times \nabla_t e_z) \tag{39}$$

and

$$e_t = -\frac{j\beta}{k_c^2}\nabla_t e_z + \frac{j\omega\mu}{k_c^2}(\hat{a}_z \times \nabla_t h_z) \tag{40}$$

where h_t and e_t are the transverse components of the magnetic and electric fields, h_z and e_z are the z components of the magnetic and electric fields, k_c is the transverse propagation constant, β is the phase constant, ε and μ are the permittivity and the permeability, and $\hat{a}_z$ is the unit vector in the z direction. For the HE_{11} mode on a stripline, consisting of a semicircular dielectric rod attached to a conducting plane, in the dielectric the transverse magnetic field is approximately parallel to the plane and the transverse

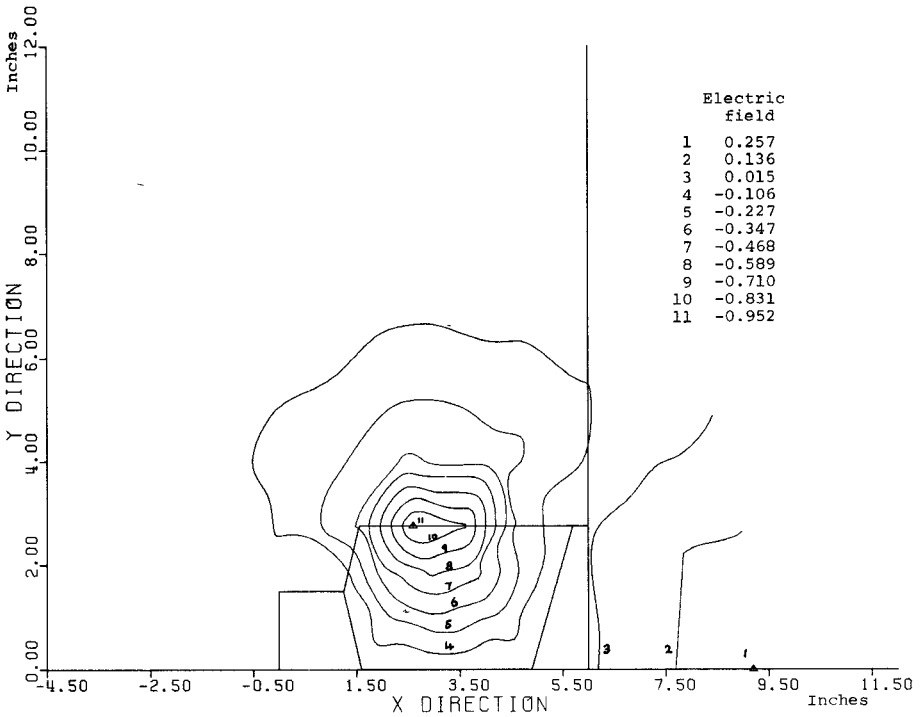

Fig. 5. Real part of electric field for an augmented rail at 1.246 GHz.

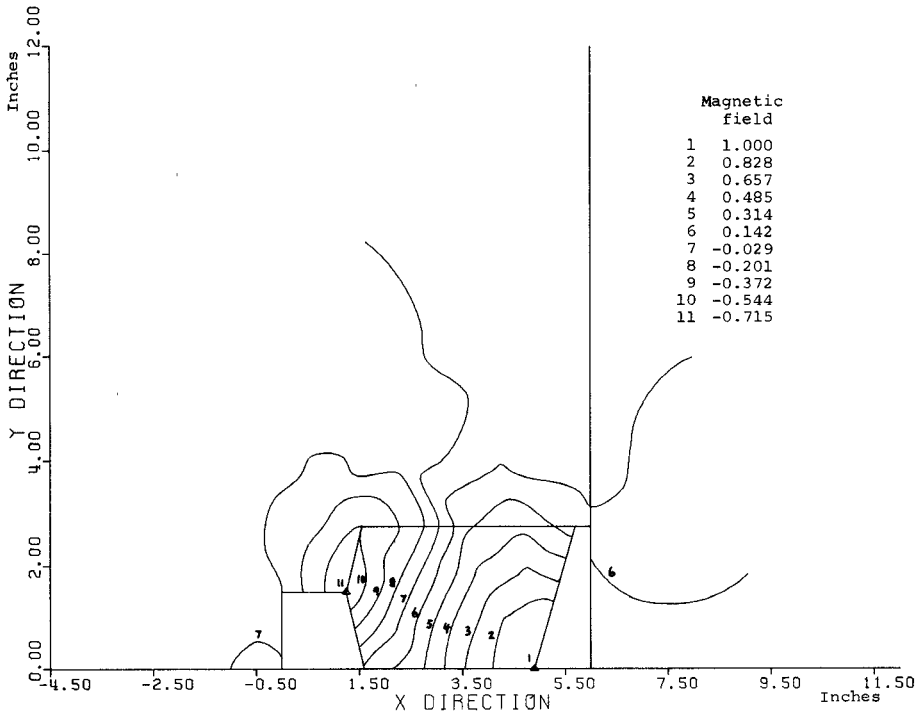

Fig. 6. Real part of magnetic field for an augmented rail at 1.246 GHz.

electric field is approximately normal to the plane. Application of (39) and (40) to Figs. 5 and 6 indicates that the transverse-field patterns follow an approximately similar shape to the previous case, if the rail contours and dielectric shape are allowed for.

Figs. 7 and 8 show, superimposed on the cross section of a rail, the real parts of the z component of the electric and magnetic fields for a frequency 0.94 GHz ($\omega = 0.5$) near the low end of the range of frequencies considered. A ratio of propagation constant to frequency of $\gamma/\omega = \beta/\omega + j\alpha/\omega = 1.07 + j0.00147$ was used to produce these figures. At this ratio of propagation constant to frequency, the imaginary part of the eigenvalue ω^2 is small relative to the values when either α/ω or β/ω are changed only slightly. The field is seen to spread much further from the rail, but is otherwise similar to the higher frequency field previously

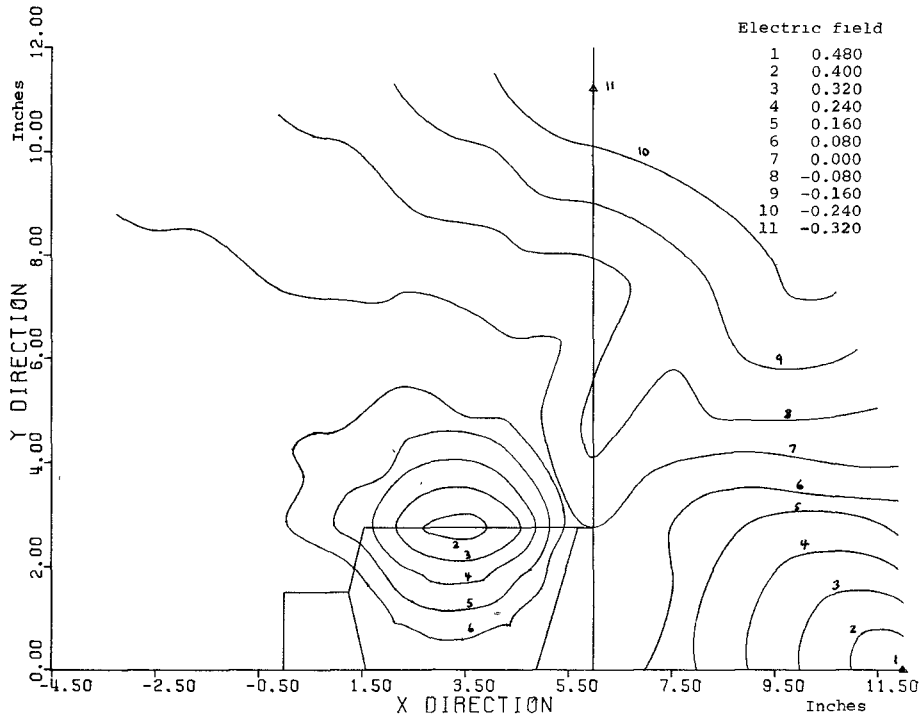

Fig. 7. Real part of electric field for an augmented rail at 0.94 GHz.

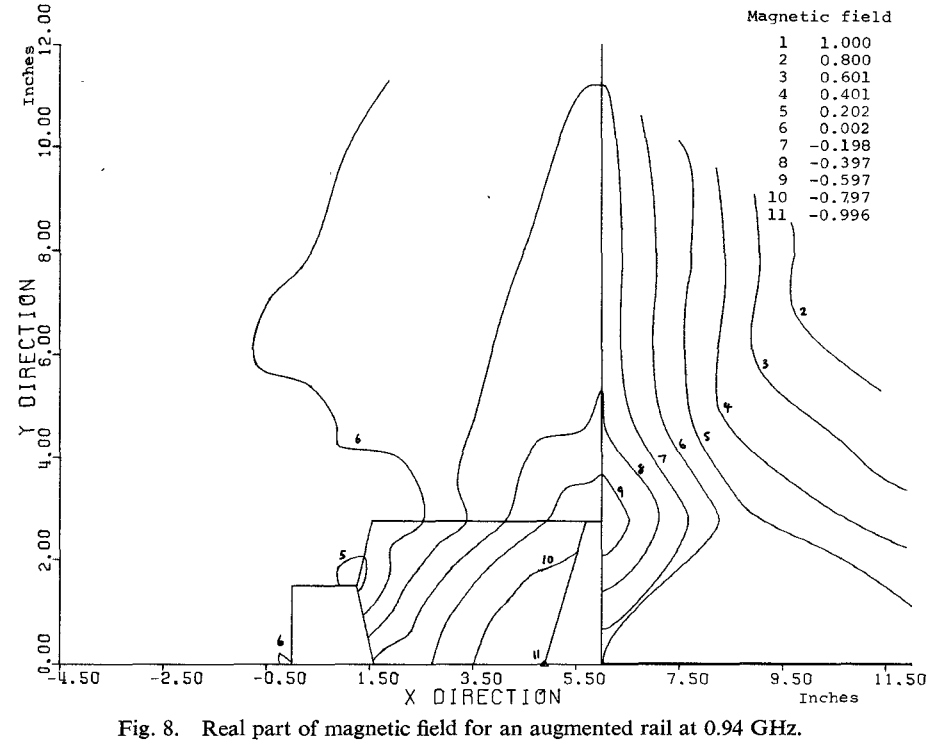

Fig. 8. Real part of magnetic field for an augmented rail at 0.94 GHz.

considered. The transverse distance from the surface of the dielectric to the point where the z component of the electric field has decayed to $1/e$ of the value at the surface is about 2.67 in as compared to 0.67 in for the midfrequency.

Figs. 9 and 10 show the field in the case where the rail is resting on rain water. The ratio of normalized attenuation to normalized frequency is $\alpha/\omega = 0.00014$ for $\beta/\omega = 1.14$, which is less than the case for dry sandy soil, $\alpha/\omega = 0.000302$. The reason may be explained by observing that, in Fig. 9, the field is largely expelled from the ground. The phase constant remains unchanged.

VI. Conclusion

A numerical procedure was developed, using the Rayleigh–Ritz variational and finite-element methods, for determining the complex propagation constants and com-

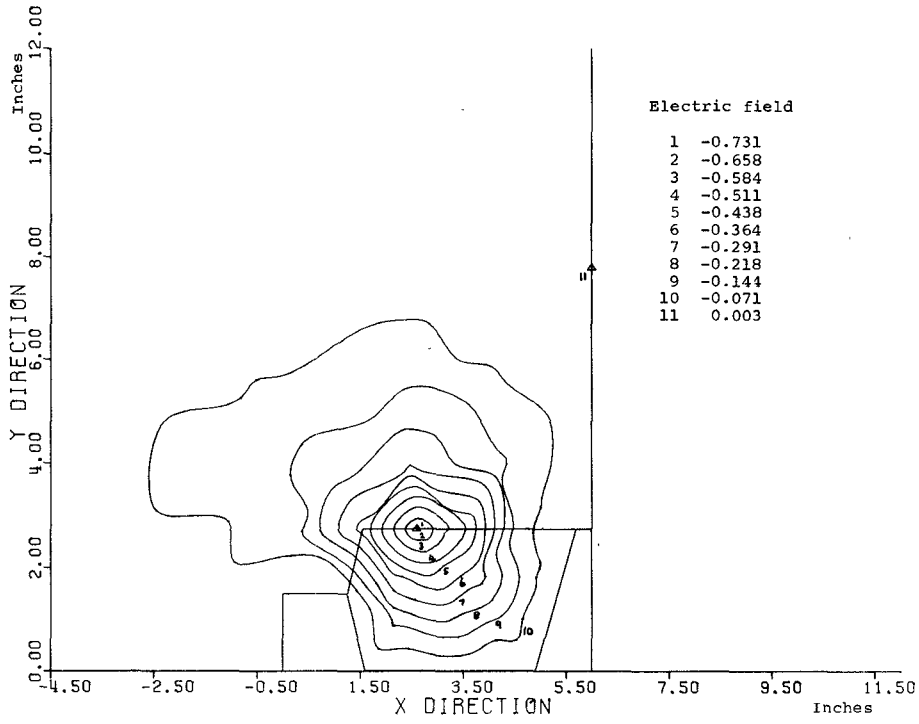

Fig. 9. Real part of electric field for an augmented rail resting on water at 1.24 GHz.

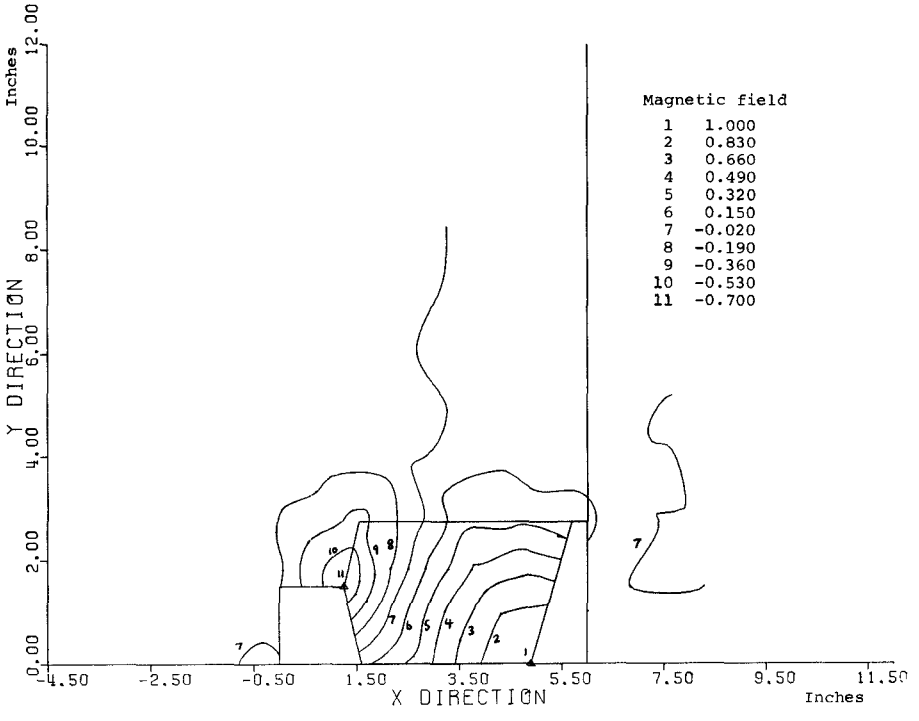

Fig. 10. Real part of magnetic field for an augmented rail resting on water at 1.24 GHz.

plex electromagnetic fields for the lowest order modes for electromagnetic waves propagating along a closed or surface waveguide. A review was presented of the results obtained in verifying this method. A railroad rail modified by the addition of a strip of dielectric material was then investigated as a waveguide, assuming typical material properties for the rail, ground, dielectric, and air. It was shown that waves of approximately 1.3 GHz may be expected to propagate a distance of 155 m before falling to $1/e$ of their amplitude at the transmitter. Water saturating the track was shown to have only a minor effect. The z component of the electric and magnetic fields was plotted for two frequencies and for both dry and water-saturated ground. The field patterns show the nature of the surface wave obtained and enable the future construction of efficient antennas for coupling into the waveguide.

The results of analyzing the railroad track as a lossy waveguide illustrate the usefulness of the numerical method developed. In addition, the proposal to use the track as a waveguide was shown to warrant further investigation as a viable alternative for collision avoidance or headway control in future rapid-transit systems.

References

[1] W. M. Elsasser, "Attenuation in a dielectric circular rod," *J. Appl. Phys.*, vol. 20, pp. 1193–1196, Dec. 1947.
[2] A. Sommerfield, *Electrodynamics*. New York: Academic, 1964, pp. 177–192.
[3] Z. J. Czendes and P. Silvester, "Numerical solutions of dielectric loaded waveguides: I—Finite element analysis," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-18, pp. 1124–1131, Dec. 1970.
[4] S. Ahmed, "Finite-element method for waveguide problems," *Electron. Lett.*, vol. 4, pp. 387–389, 1968.
[5] P. Daly, "Hybrid mode analysis of microstrip by finite element methods," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-19, pp. 19–25, Jan. 1971.
[6] B. H. McDonald and A. Wexler, "Finite-element solution of unbounded field problems," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, pp. 841–847, Dec. 1972.
[7] S. Ahmed and P. Daly, "Waveguide solutions by the finite element method," *Radio Electron. Eng.*, vol. 38, p. 217, Oct. 1969.
[8] ——, "Finite element methods for inhomogeneous waveguides," *Proc. IEE*, vol. 116, p. 1661, Oct. 1969.
[9] P. L. Arlett *et al.*, "Application of finite elements to the solution of Helmholtz's equation," *Proc. IEE*, vol. 115, p. 1962, Dec. 1968.
[10] P. Daly, "Finite elements for field problems in cylindrical coordinates," *Int. J. Numer. Methods Eng.*, vol. 6, pp. 167–168, 1973.
[11] R. E. Collins, *Field Theory of Guided Waves*. New York: McGraw-Hill, 1960, p. 454.
[12] B. A. Finlayson, *Methods of Weighted Residuals and Variational Principles with Application in Fluid Mechanics, Heat and Mass Transfer*. New York: Academic, 1972, pp. 7–12, 299–232.
[13] A. D. McAulay, "Detection of track guided ground vehicles using the track as an electromagnetic waveguide," Ph.D. Thesis, Carnegie-Mellon Univ., 1974.
[14] ——, "The finite element solution of dissipative electromagnetic surface waveguides," *Int. J. Numer. Methods Eng.*, vol. 11, pp. 11–25, 1977.
[15] ——, "Track guided radar for rapid transit systems," *AIAA J. Aicraft*, vol. 12, pp. 676–681, Sept. 1975.
[16] ——, "Progress in signaling for track guided systems," *Transportation Eng. J. ASCE*, vol. 101, no. TE4, pp. 621–637, Nov. 1975.
[17] O. C. Zienkiewitz and Y. K. Cheung, *The Finite Element Method in Structural and Continuum Mechanics*. New York: McGraw-Hill, 1970.
[18] J. T. Oden, *Finite Element of Non-Linear Continua*. New York: McGraw-Hill, 1972.
[19] S. G. Miklin and K. L. Smolitsky, *Approximate Methods for the Solution of Differential and Integral Equations*. New York: Elsevier, 1967, ch. 3.
[20] D. J. Richards and A. Wexler, "Finite element solutions within curved boundaries," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, p. 650, Oct. 1972.
[21] P. Silvester, "High order polynomial triangular finite elements for potential problems," *Int. J. Eng. Sci.*, vol. 7, p. 849, Aug. 1969.
[22] P. Daly, "Finite element coupling matrices," *Electron. Lett.*, vol. 5, p. 613, Nov. 1969.
[23] C. B Moler and G. W. Stewart, "An algorithm for the generalized matrix eigenvalue problem $Ax = \lambda Bx$," CS-232-71, Stanford Univ.
[24] G. Goubau, "Surface waves and their application to transmission lines," *J. Appl. Phys.*, vol. 21, pp. 1119–1128, Nov. 1950.
[25] H. M. Barlow and J. Brown, *Radio Surface Waves*. Oxford: Oxford University Press, 1962.
[26] *Reference Data for Radio Engineers*, ITT, 5th ed., 1969.
[27] A. S. Hu, "Transmission properties of waveguide communication systems," *Proc. IEEE*, vol. 61, pp. 556–561, May 1973.
[28] H. R. Reed and C. M. Russell, *Ultra-High Frequency Propagation*. New York: Wiley, 1953, p. 24.

Aperture Coupling Between Microstrip and Resonant Cavities

DAVID S. JAMES, MEMBER, IEEE, GUY R. PAINCHAUD, MEMBER, IEEE, AND WOLFGANG J. R. HOEFER, MEMBER, IEEE

Abstract—**This paper presents a simple analysis for the coupling between microstrip and a cavity through an aperture located in the substrate ground plane. The analysis is based on Wheeler's equivalent-energy concept for small-hole coupling and an approximate parallel-plate waveguide model for the microstrip. The theory appears adequate for most design purposes, and has been used successfully in the design of stabilizing cavities for experimental 12-GHz low-noise FET oscillators.**

I. Introduction

THE PURPOSE of this paper is to present an analysis and experimental data on the performance of a novel microstrip-to-cavity transition, so that high Q cavities can be made compatible with existing MIC techniques. Previous authors [1]–[5] have described empirically designed microstrip and stripline-to-waveguide transitions; however, with the exception of [4], none of these papers describe coupling to a resonant cavity.

The transition studied is shown in Fig. 1. The ground plane of the microstrip substrate forms one of the cavity end plates. In particular, we have investigated coupling to cylindrical cavities resonating in the TE_{01n} mode. Coupling is by means of an aperture located in the ground plane at the point of maximum radial H-field in the cavity. The microstrip line terminates in an open circuit $3\lambda_g/4$ beyond the aperture. This length of line maximizes the coupling through the aperture [6]–[8]. This configuration is quite practical, as the cavity can be machined into the substrate holder.

Unfilled cavities can readily yield values of unloaded $Q(Q_0)$ of at least 25 000. This value is much greater than that of both planar resonators ($Q_0 < 500$) [9] and "open"

Manuscript received December 10, 1975; revised August 9, 1976.

D. S. James was with the Department of Communications, Communications Research Centre, Ottawa, Ont., Canada. He is now with the Ferranti Solid State Microwave Group, Manchester, U.K.

G. R. Painchaud was with the Department of Electrical Engineering, University of Ottawa, Ottawa, Ont., Canada. He is now with the Department of Communications, Communications Research Centre, Ottawa, Ont., Canada.

W. J. R. Hoefer is with the Space Division, AEG-Telefunken, Backnang, Germany, on leave from the Department of Electrical Engineering, University of Ottawa, Ottawa, Ont., Canada.