(4)

THE OHIO STATE UNIVERSITY

JOINT SERVICES ELECTRONICS PROGRAM

Fifteenth Annual Report Appendix

The Ohio State University

ElectroScience Laboratory

Department of Electrical Engineering
Columbus, Ohio 43212

Annual Report Appendix 721563-5
Contract No. N00014-89-J-1007
November 1992

Department of the Navy
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO.	2.	3. Recipient's Accession No.
4. Title and Subtitle Appendix Joint Services Electronic Program – Fifteenth Annual Report			5. Report Date November 1992 6.
7. Author(s) Leon Peters, Jr.			8. Performing Org. Rept. No. 721563-5
9. Performing Organization Name and Address The Ohio State University ElectroScience Laboratory 1320 Kinnear Road Columbus, OH 43212			10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No. (C) N00014-89-J-1007 (G)
12. Sponsoring Organization Name and Address Department of the Navy, Office of Naval Research 800 North Quincy Street Arlington, Virginia 22217			13. Report Type/Period Covered Annual Report Appendix 14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

17. Document Analysis a. Descriptors

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
A. Approved for public release; Distribution is unlimited.	Unclassified	
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18) *See Instructions on Reverse* OPTIONAL FORM 272 (4-77)
Department of Commerce

Contents

INTRODUCTION

This Appendix contains the reprints published under JSEP in the time September 1991 to September 1992.

In addition to the reprints contained herein, there are 7 papers already accepted for publication during the next contract period, 11 papers submitted and 11 papers in preparation.

JSEP REFEREED JOURNAL PAPERS PUBLISHED SEPTEMBER 1991 TO SEPTEMBER 1992

1. E.H. Newman and K. Kingsley, "An Introduction to the Method of Moments," *Computer Physics Communications*, vol. 68, 1991, pp. 1–18, (invited paper).

2. R.J. Burkholder, R.-C. Chou and P.H. Pathak, "Two Ray Shooting Methods for Computing the EM Scattering by Large Open-Ended Cavities," *Computer Physics Communications*, 1991, pp. 353–365, (invited paper).

3. R.J. Burkholder and P.H. Pathak, "Analysis of EM Penetration into and Scattering by Electrically Large Open Waveguide Cavities Using Gaussian Beam Shooting," *Proceedings of the IEEE*, vol. 79, no. 10, October 1991, pp. 1401–1412.

4. J. Li and R.T. Compton, Jr., "Angle Estimation using a Polarization Sensitive Array," *IEEE Transactions on Antennas and Propagation*, vol. 39, October 1991, pp. 1539–1543.

5. M.S. Kluskens and E.H. Newman, "A Microstrip Line on a Chiral Substrate," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, November 1991, pp. 1889–1991.

6. M.S. Kluskens and E.H. Newman, "Scattering by a Chiral Cylinder of Arbitrary Cross Section in the Presence of a Half-Plane," *Journal Electromagnetic Waves and Applications*, vol. 6, 1992, pp. 721–731.

7. P.H. Pathak, "High Frequency Techniques for Antenna Analysis," *Proceedings of IEEE*, vol. 80, no. 1, January 1992, pp. 44–65, (invited paper).

8. S. Barkeshli and P.H. Pathak, "On the Dyadic Green's Function for a Planar Multilayered Dielectric/Magnetic Media," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 1, January 1992, pp. 128–142.

9. J. Ward and R.T. Compton, Jr., "Improving the Performance of a Slotted ALOHA Packet Radio Network with an Adaptive Array," *IEEE Transactions on Communications*, vol. 40, no. 2, February 1992, pp. 292–300.

10. R. Lee and A.C. Cangellaris, "A Study of Discretization Error in the Finite Element Approximation of Wave Solutions," *IEEE Transactions on Antennas and Propagation*, vol. 40, May 1992, pp. 542–549.

11. J. Li and R.T. Compton, Jr., "Two Dimensional Angle and Polarization Estimation Using the ESPRIT Algorithm," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 5, May 1992, pp. 550–555.

12. G.A. Somers and P.H. Pathak, "Uniform GTD Solution for the Diffraction by Metallic Tapes on Panelled Compact-Range Reflectors," *IEE Proceedings-H*, vol. 139, no. 3, June 1992, pp. 297–305.

13. R.G. Rojas, "Integral Equations for the Scattering by a Three Dimensional Inhomogeneous Chiral Body," *Journal Electromagnetic Waves and Applications*, vol. 6, no. 5/6, July 1992, pp. 733–750.

14. F.W. Vook and R.T. Compton, Jr., "Bandwidth Performance of Linear Adaptive Arrays with Tapped Delay-Line Processing," (Correspondence), *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, July 1992, pp. 901–908.

The following reprint from 1991 was not included in the 1991 Annual Appendix.

1. G. Pelosi, R. Tiberio and R.G. Rojas, "Electromagnetic Field Excited by a Line Source Placed at the Edge of an Impedance Wedge," *IEEE Transactions on Antennas and Propagation*, vol. 39, July 1991, pp. 1043–1046.

JSEP RELATED REFEREED JOURNAL PAPERS ACCEPTED FOR PUBLICATION SEPTEMBER 1991 TO SEPTEMBER 1992

1. P.H. Pathak and R.J. Burkholder, "A Reciprocity Formulation for Calculating the EM Scattering by an Obstacle within an Open-Ended Waveguide Cavity," *IEEE Transactions on Microwave Theory and Techniques.*

2. J. Li and R.T. Compton, Jr., "Maximum Likelihood Angle Estimation for Signals with Known Waveforms," *IEEE Transactions on Signal Processing.*

3. R.G. Rojas and M. Otero, "Scattering by a Resistive Strip Attached to an Impedance Wedge," *Journal of Electromagnetic Waves and Applications.*

4. J. Li and R.T. Compton, Jr., "Angle and Polarization Estimation in a Coherent Signal Environment," *IEEE Transactions on Aerospace and Electronic Systems*, July 1993.

5. J. Ward and R.T. Compton, Jr., "High Throughput Slotted ALOHA Packet Radio Networks with Adaptive Arrays," *IEEE Transactions on Communications.*

6. M. Marin and P.H. Pathak, "An Asymptotic Closed-Form Representation for the Grounded Double Layer Surface Green's Function," *IEEE Transactions on Antennas and Propagation.*

7. H.C. Ly and R.G. Rojas, "Analysis of Diffraction by Material Discontinuities in Thin Material Coated Planar Surfaces based on Maliuzhnets' Method," *Radio Science.*

JSEP RELATED PAPERS SUBMITTED FOR PUBLICATION SEPTEMBER 1991 TO SEPTEMBER 1992

1. F.W. Vook and R.T. Compton, Jr., "Adaptive Array Beamforming in a Packet Radio Network," *IEEE Transactions on Communications.*

2. P. Munk and P. Pathak, "Analysis of EM Scattering by an Array of Waveguide Fed Slots in a Dielectric Filled Rectangular Cavity Opening into a Ground of a Plane, *Radio Science.*

3. H.C. Ly, R.G. Rojas and P.H. Pathak, "EM Plane Wave Diffraction by a Planar Juntion of Two Thin Material Half-Planes — Oblique Incidence," *IEEE Transactions on Antennas and Propagation.*

4. R.G. Rojas and L.M. Chou, "Generalized Impedance/Resistive Boundary Conditions for a Planar Chiral Slab," *Radio Science.*

5. J.L. Blanchard and E.H. Newman, "Integral Equation Analysis of Artificial Media," *IEEE Transactions on Antennas and Propagation.*

6. M.E. Peters and E.H. Newman, "Analysis of an Artificial Dielectric Composed of Small Dielectric Spheres," *IEEE Transactions on Antennas and Propagation.*

7. R. Torres and E.H. Newman, "Integral Equation Analysis of a Sheet Impedance Coated Window Slot Antenna," *IEEE Transactions on Antennas and Propagation.*

8. U. Pekel and R. Lee, "An A Posteriori Error Reduction Scheme for the Three Dimensional Finite Element Solution of Maxwell's Equations," *IEEE Transactions on Microwave Theory and Techniques.*

9. N. Wang and L. Peters, Jr., "Scattering by Thin Wire Loaded with a Ferrite Ring," *IEEE Transactions on Antennas and Propagation.*

10. K.C. Hill and P.H. Pathak, "On the Nature and Evaluation of the Transition Function for a UTD Corner Diffraction Coefficient," *ACES Journal.*

11. Y.S. Choi-Grogan, R. Lee and R.-C. Chou, "An Analysis of the Effects of Boundary Conditions on Discretization Error in the Helmholtz Equation," submitted to *IEEE Transactions on Antennas and Propagation.*

JSEP RELATED PAPERS IN PREPARATION FOR PUBLICATION SEPTEMBER 1991 TO SEPTEMBER 1992

1. P.H. Pathak, A. Nagamune and R.G. Kouyoumjian, "An Analysis of Compact Range Measurements."

2. P.H. Pathak, P. Law and R.J. Burkholder, "High Frequency Electromagnetic Scattering by a Large Obstacle/Termination within an Open Cavity Structure."

3. M. Hsu, P.H. Pathak and C.W. Chuang, "Analysis of the Asymptotic HF EM Coupling Between Sources Anywhere in the Vicinity of a Circular Cylinder."

4. K.C. Hill and P.H. Pathak, "A Uniform Stationary Phase Evaluation of a Double Integral with Algebraic Singularities."

5. K.C. Hill and P.H. Pathak, "An Approximate UTD Corner Diffraction Coefficient."

6. H.T. Anastassiu and P.H. Pathak, "High Frequency Analysis of Gaussian Beam Scattering by a Two-Dimensional Parabolic Contour of Finite Width."

7. L.M. Chou and R.G. Rojas, "A WH/GSMT Based Full-Wave Analysis of Multilayered Printed Transmission Lines." to be submitted to *IEEE Transactions on Microwave Theory and Techniques.*

8. L.M. Chou, R.G. Rojas and P.H. Pathak, "Dispersion and Lateral Leakage of Conductor Backed Coplanar Waveguide with Layered Substrate and Finite-Extent Ground Planes," to be submitted to *IEEE Transactions on Microwave Theory and Techniques.*

9. R. Lee and V. Chupongstimun, "A Partitioning Scheme for Electromagnetic Scattering from Electrically Large Cylinders."

10. J. O. Jevtic and R. Lee, "Dispersion Analysis of the Wave Equation for Two and Three Dimensional Edge Elements."

11. U. Pekel and R. Lee, "A Three-Dimensional Finite Element Method for Electromagnetic Scattering from Objects in an Unbounded Region."

JSEP RELATED CONFERENCES/ORAL PRESENTATIONS SEPTEMBER 1991 TO SEPTEMBER 1992

1. J. Li and R.T. Compton, Jr., "Maximum Likelihood Angle Estimation for Signals with Known Waveforms," IEEE Signal Processing Society, Seventh Workshop on Multidimensional Signal Processing, Lake Placid, New York, September 23–25, 1991.

2. R.T. Compton, Jr., "Adaptive Antennas in Packet Radio," 25th Annual Asilomar Conference on Signals, Systems and Computers, November 4–6, 1991, Pacific Grove, California.

3. J. Li and R.T. Compton, Jr., "Performance Analysis for Angle and Polarization Estimation using ESPRIT," 1992 International Conference on Acoustics, Speech and Signal Processing, San Francisco, California, March 23–26, 1992.

4. L.M. Chou, R.G. Rojas and P.H. Pathak, "A WH/GSMT Based Full-Wave Analysis of the Power Leakage from Conductor-Backed Coplanar Waveguides," 1992 MTT International Symposium, Albequerque, New Mexico, June 1992.

5. L.M. Chou, R.G. Rojas and P.H. Pathak, "A WH/GSMT Based Full-Wave Analysis of Multilayered Printed Transmission Lines," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

6. H.C. Ly, R.G. Rojas and P.H. Pathak, "Diffraction from a Two-Part Planar Material Junction: Oblique Incidence Case," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

7. R.G. Rojas and M.F. Otero, "EM Scattering by a Chiro-Dielectric Body of Arbitrary Shape in the Presence of an Impedance Wedge," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

8. H.C. Ly and R.G. Rojas, "Application of the Maliuzhnets' Method to Diffraction Problems involving Generalized Impedance Boundary Conditions," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

9. U. Pekel and R. Lee, "An Element-by-Element A Posteriori Error Estimation and Improvement Approach for the Finite Element Analysis of Three-Dimensional Electromagnetic Boundary Value Problems," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

10. Y. S. Choi-Grogan, R. Lee, and R. C. Chou, "Discretization Error for FEM," Joint AP-S Symposium and URSI Radio Science Meeting, Chicago, IL, July, 1992.

11. M. Hsu, C.W. Chuang, P.H. Pathak and R.-C. Chou, "An Asymptotic Analysis of the Near Field EM Scattering from a Smooth Convex Body," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

12. L. Peters, Jr., "Historical Review of ElectroScience Laboratory with an Emphasis on Contributions of R.G. Kouyoumjian," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992 (invited).

13. R.G. Kouyoumjian, "A Brief History of the UTD — Focus on Edge Diffraction," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992 (invited).

14. P.H. Pathak and R.J. Burkholder, "On the Question of Time Causality for HF Ray Fields Traversing Caustics," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992 (invited).

15. P.H. Pathak and R.J. Marhefka, "On the Behaviour of Uniform Ray Solutions at Lower Frequencies," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992 (invited).

16. G.A. Somers and P.H. Pathak, "An Efficient Analysis of the Mutual Coupling in a Large Finite Array of Slots in a Material Coated Ground Plane," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

17. R-C. Chou, T.T. Chia and R. Lee, "The Energy Flow Inside a Waveguide Cavity using the SBR and GRE Methods," 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992.

18. SHORT COURSE:
P.H. Pathak and Y. Rahmat-Samii, Short Course on Asymptotic High Frequency, 1992 IEEE APS/URSI International Symposium, Chicago, Illinois, July 1992 (invited).

JSEP RELATED M.SC. THESES AND PH.D. DISSERTATIONS SEPTEMBER 1991 TO SEPTEMBER 1992

Dissertations:

1. H.C. Ly, "A UTD Analysis of EM Diffraction by an Abrupt Discontinuity in Thin Planar Material Configurations," Ph.D. Dissertation, Department of Electrical Engineering, Ohio State University, Columbus, Ohio, March 1992.

2. F.W. Vook, "Signal Processing Arrays in Packet Radio Networks," Ph.D. Dissertation, Department of Electrical Engineering, Ohio State University, Columbus, Ohio, August 1992.

Thesis:

1. H. Anastassiu, "High Frequency Analysis of Gaussian Beam Scattering by a Parabolic Surface containing an Edge," M.Sc. Thesis, Department of Electrical Engineering, Ohio State University, Columbus, Ohio, May 1992.

2. P.R. Rousseau, "A Study of Electromagnetic Scattering by Obstacles Placed within Open Cavities," M.Sc. Thesis, Department of Electrical Engineering, Ohio State University, Columbus, Ohio, August 1992.

REPRINTS
JSEP REFEREED JOURNAL PAPERS
SEPTEMBER 1991 TO SEPTEMBER 1992

Computer Physics Communications 68 (1991) 1–18
North-Holland

Computer Physics Communications

An introduction to the method of moments *

E.H. Newman and K. Kingsley

The Ohio State University, Department of Electrical Engineering, ElectroScience Laboratory, 1320 Kinnear Rd., Columbus, OH 43212, USA

Received 31 August 1990; in revised form 20 December 1990

This paper will present an introduction to the theory and application of the method of moments (MM) to problems of electromagnetic radiation and scattering. The MM procedure for solving a linear operator equation, by transforming it into a matrix equation, is reviewed. The integral equation and MM solution for an arbitrary perfectly conducting body is presented, and then illustrated by the simple example of scattering by a 2D perfectly conducting strip. Numerical results are used to illustrate the accuracy, convergence, and typical computer CPU times. The MM solution for radiation or scattering by a rectangular dielectric cylinder is also presented. Finally, the internal resonance problem is discussed.

1. Introduction

This paper will present an introduction to a numerical technique known as the *moment method* or *method of moments* (MM), especially as it applies to problems in electromagnetic radiation and scattering [1,2]. The MM is a numerical technique for solving a linear operator equation by transforming it into a system of simultaneous linear algebraic equations, i.e. a matrix equation. Over the last 30 years the MM has been extensively applied to virtually every area of electromagnetics including radiation and/or scattering by perfectly conducting and material bodies, thin wire antennas, aperture penetration, printed circuit structures, etc [3]. This paper will describe some of the basic features of the MM, rather than present an extensive bibliography of MM research and applications.

In electromagnetics a "moment method solution" usually refers to a problem in which the MM is used to solve a linear integral equation for a current distribution representing a body. This is basically a two-step procedure: The first step is to obtain the integral equation. Basically, the surface or volume equivalence theorems are used to replace the body by free space and by equivalent currents [4]. The use of these equivalence theorems is crucial in that it allows the integral equation and MM solution to be formulated entirely in terms of the *free space* fields of currents. The integral equation for the equivalent currents is obtained either as a statement of the boundary conditions of the problem, or as a statement of the equivalence theorems used. The second step is to solve the integral equation by the MM. In brief, the unknown current is expanded in terms of an appropriate set of basis functions. If N terms are retained in the expansion for the current, then N weighted averages of the integral equation are enforced. This MM procedure transforms the integral equation into an order N matrix equation for the N coefficients in the expansion for the current. Once the current is known, most parameters of engineering interest such as input impedance or radiated fields, can be found in a straightforward manner, since they involve only the *free space* fields of the currents.

The main advantage of MM solutions is high accuracy. The MM is a direct numerical solution

* This work was sponsored by the Joint Services Electronics Program under Contract N00014-78-C-0049 with the The Ohio State University Research Foundation.

of the (usually) exact integral equation. All phenomena of the problem are inherent in the integral equation, and are thus automatically included in the MM solution. For example, MM solutions in electromagnetics automatically include surface waves, creeping waves, multiple diffractions, shadowing effects, etc. A second advantage of the MM is that it is capable of dealing with very complex geometries. In fact, several user-oriented computer codes have been written which can treat geometries as simple as a dipole or as complex as an airplane [5–10].

The main limitation of the MM is a result of the fact that N, the number of terms which must be retained in the expansion for the current in order to obtain reasonable accuracy, is proportional to the electrical size of the body. The CPU time to set up and store the MM matrix equation is proportional to N^2. The CPU time to solve the MM matrix equation is proportional to N^3 [11]. Thus, as the frequency is increased, the required computer power also increases, and at some point it becomes so large that the MM solution is impractical. However, it should be emphasized that this is a machine or hardware limitation, and that as computers become more powerful, it will be practical to apply the MM to larger problems.

Consider the example where the unknown is the current on the surface of a 3D body. Accounting for both polarizations of the vector current, typically N will be on the order of 100 unknowns per square wavelength of surface area. Thus, treating a body of $10\lambda^2$ would require dealing with an order $N = 1000$ matrix equation. Setting up and solving the MM matrix equation typically will require about an hour of CPU time. Storage of the MM matrix equation will involve more than $N^2 = 1$ million complex numbers. Note that doubling the frequency will increase the surface area in λ^2 and thus N by a factor of 4. This will increase the required storage by a factor of 16, the CPU time to set up the MM matrix equation by a factor of 16, and the CPU time to solve the matrix equation by a factor of 64. Thus, the MM is often referred to as a low frequency technique, applicable when the body is not electrically large.

Section 2 of this paper describes the MM procedure, and discusses the important question of convergence. Section 3 describes the MM solution for an arbitrary perfectly conducting body, and then illustrates the method by the simple example of scattering by a perfectly conducting cylinder. Numerical data will illustrate the accuracy, convergence, and CPU times for the MM solution. Section 4 presents the MM solution for radiation or scattering by a dielectric body. Finally, section 5 will briefly describe the internal resonance problem.

In this paper all electromagnetic fields and currents are considered to be time harmonic, with the $e^{+j\omega t}$ time dependence suppressed. Also, all CPU times are for a VAX 8550 computer which is about six times faster than a VAX 11/780.

2. The method of moments

2.1. The MM procedure

This section will describe the mathematical procedure known as the method of moments (MM), that is the procedure for solving a linear operator equation by transforming it into a system of simultaneous linear algebraic equations, commonly referred to as a matrix equation. The description given here for the MM largely follows that originally presented by Harrington [1,2].

An inhomogeneous linear operator equation can be written as

$$L(f) = g, \tag{1}$$

where L is a known linear operator, g is a known source or excitation function, and f is the unknown response function. In electromagnetics f is typically an equivalent current which produces the radiation or scattering from the body, $L(f)$ is typically the electromagnetic field of the current f, and g is a known incident field. The linear operator equation is typically an expression of either the boundary conditions of the problem or of the equivalence theorems [4] used to define the current f.

The first step in the MM solution of eq. (1) is to expand the unknown function f as

$$f \approx f^N = \sum_{n=1}^{N} a_n f_n, \tag{2}$$

where f^N is an N term expansion or approximation of f, the f_n are a series of N known linearly independent expansion or basis functions in the domain of L, and the a_n are a series of N unknown constants to be determined by the MM. Substituting the f^N expansion of eq. (2) into the original operator (1), and using the linearity of L yields

$$Lf^N = \sum_{n=1}^{N} a_n L(f_n) = g. \tag{3}$$

For finite N, f^N in eq. (2) will generally not be a complete expansion for f, and thus in practice it is not possible to choose the a_n to exactly satisfy eqs. (2) or (3). Instead, the a_n are chosen so that N weighted averages of eq. (3) are satisfied. Define a series of N linearly independent weighting or testing functions in the range of L, denoted w_m, $m = 1, 2, \ldots, N$. If the weighting functions are chosen identical to the expansion functions, i.e. $w_m = f_m$, then the MM solution is referred to as *Galerkin's method*. The bracket notation $\langle u, w_m \rangle$ will be used to denote the inner product between the functions u and w_m. Typically, the inner product is taken as

$$\langle u, w_m \rangle = \int_m u w_m^* \, \mathrm{d}r, \tag{4}$$

where the integral is over the region (line, surface, or volume) of w_m and the * implies complex conjugate. In Harrington's original description of the MM the weighting function was not conjugated. Although the choice is arbitrary, here we include the complex conjugate so that we can directly use the results of recent work concerning the convergence of Galerkin's method.

Taking the inner product of both sides of eq. (3) with each w_m and using the linearity of L results in

$$\sum_{n=1}^{N} a_n \langle Lf_n, w_m \rangle = \langle g, w_m \rangle, \quad m = 1, 2, \ldots, N. \tag{5}$$

Equation (5) can be written as

$$\langle Lf^N - g, w_m \rangle = 0, \quad m = 1, 2, \ldots, N. \tag{6}$$

Thus, the MM determines the a_n so that the difference between Lf^N and g is orthogonal to the N weighting functions.

Equation (5) can be recognized as N simultaneous linear algebraic equations in the N unknowns a_n, $n = 1, 2, \ldots, N$. It can be more compactly written in matrix form as

$$[L]A = G, \tag{7}$$

where $[L]$ is an $N \times N$ coefficient matrix, G is the length N right-hand-side vector, and A is the length N solution vector which contains the a_n from eq. (2). Typical elements of $[L]$ and G are given by

$$L_{mn} = \langle Lf_n, w_m \rangle, \quad m, n = 1, 2, \ldots, N, \tag{8}$$

$$G_m = \langle g, w_m \rangle, \quad m = 1, 2, \ldots, N. \tag{9}$$

Using standard matrix algebra, eq. (7) can now be solved for the solution vector A, which when substituted into eq. (2) provides an approximation to f. In electromagnetics, Galerkin's method results in a symmetric $[L]$ matrix. Note that the matrix $[L]$ is independent of the excitation function g. Thus, an advantage of MM solutions is that a relatively small effort is required to obtain the solution for a second or subsequent excitation, since on the first solution one must set up and LU decompose $[L]$.

2.2. Convergence of the MM

One of the most important questions concerning the MM is that of convergence. That is, as $N \to \infty$, does $f^N \to f$? Very little can be said concerning the convergence of the general MM in which weighting functions can be chosen differently from the expansion functions, and thus

the remarks in this section will refer to Galerkin's method in which $w_m = f_m$.

The convergence of Galerkin's method is dependent upon the properties of the operator L [12,13]. If L is positive, i.e. if $\langle Lf, f\rangle > 0$ for all $f \neq 0$, then it can be shown that f^N converges in energy to f [13], i.e.

$$\lim_{N\to\infty} \sqrt{\langle L(f^N - f), (f^N - f)\rangle} = 0. \tag{10}$$

Note that convergence in energy does not imply that f^N converges to f.

In electromagnetics the L operators are not positive, and thus nothing can be said concerning the convergence of Galerkin's method or the MM. However, experience has shown that with a "reasonable" choice of expansion and weighting functions the MM does converge, and in fact is often used as a reference solution and referred to as an exact method. Unfortunately, the choice of "reasonable" expansion and weighting functions remains somewhat of an art. Basically, one wishes to choose expansion functions which:

1. incorporate as many of the known properties of the unknown function f as is possible. For example if f is continuous, then it is desirable to choose expansion functions which are continuous. If f is zero at the boundaries, then it is desirable to choose expansion functions which are zero at the boundaries. If f is not zero at the boundaries, then using expansion functions which are zero at the boundaries would be considered an "unreasonable" choice, and the MM would not be expected to converge.

2. permit the inner products in eqs. (8) and (9) to be evaluated with reasonable ease.

In practice the choice of basis functions is a compromise between the above two criteria. Choosing very simple basis functions often reduces the difficulty in developing the MM code, however, the code may be slowly converging and require a relatively large N to achieve accurate results. On the other hand, one can choose very sophisticated basis functions which yield accurate results with a relatively small N. However, this can increase both the man hours required to develop the code and also the CPU time to evaluate the inner products.

2.3. The method of least squares

This section will show that the original operator equation can be modified so that the linear operator is positive, and thus convergence in energy is guaranteed. The method employs the adjoint operator, L^a, defined by

$$\langle Lu, v\rangle = \langle u, L^a v\rangle. \tag{11}$$

Operation by L^a on both sides of eq. (1) yields the new operator equation

$$L^a L(f) = L^a(g). \tag{12}$$

$L^a L$ is a positive operator since

$$\langle L^a Lu, u\rangle = \langle Lu, Lu\rangle = \| Lu \|^2 > 0, \quad \text{if} \quad u \neq 0, \tag{13}$$

and thus the Galerkin Method solution of eq. (12) will converge in energy. The elements in the matrix equation (7) are given by

$$L_{mn} = \langle Lf_n, Lf_m\rangle, \quad m, n = 1, 2, \ldots, N, \tag{14}$$

$$G_m = \langle g, Lf_m\rangle, \quad m = 1, 2, \ldots, N. \tag{15}$$

Equations (14) and (15) can be recognized as the *method of least squares* which can be derived by direct minimization of $\| Lf^N - g \|$ [14]. Comparison with eqs. (8) and (9) shows that the method of least squares is equivalent to a MM solution of the original operator (1), but with weighting functions chosen as $w_m = Lf_m$. The main disadvantage of the method of least squares is that the matrix elements are typically more difficult and time consuming to evaluate.

Figure 1 shows a comparison of the mean square error, $\| \epsilon_N \|^2$, versus N in the solution of a simple differential equation by Galerkin's method and by the method of least squares [15]. Note that the error in the least squares solution monotonically decreases with increasing N, while that for Galerkin's method does not. Also, the error for the least squares solution is always lower than that for Galerkin's method, however, the difference is small. In electromagnetics the

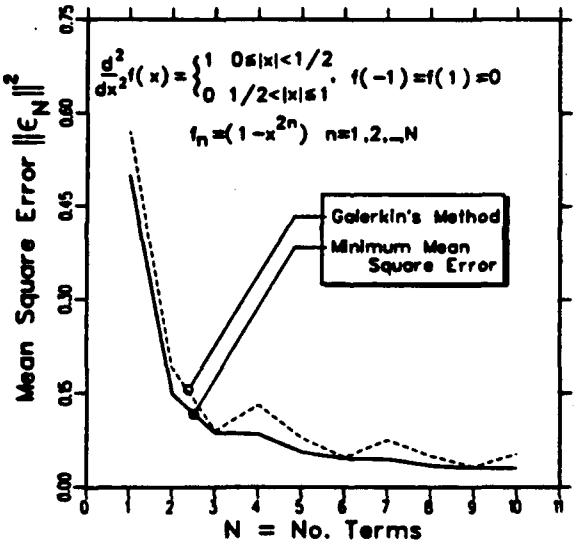


Fig. 1. A comparison of the mean square error in a Galerkin method and a least square error solution of a differential equation.

method of least squares is employed far less than the MM, probably because the advantage in the increased rate of convergence does not justify the increased computational effort.

3. Perfectly conducting bodies

This section will outline the integral equation and MM solution to the problem of radiation or scattering by a perfectly conducting body. First, the solution for a general perfectly conducting body will be presented, and then the method will be illustrated by the simple problem of plane wave scattering by a perfectly conducting cylinder.

3.1. The integral equation

This section will obtain the electric and magnetic field integral equations for the current distribution on a perfectly conducting surface. The general problem is illustrated in fig. 2a where the assumed known impressed currents $(\boldsymbol{J}^i, \boldsymbol{M}^i)$ radiate the unknown total fields $(\boldsymbol{E}, \boldsymbol{H})$ in a medium which is free space except for the presence of a perfectly conducting body enclosed by the surface S. Here (μ_0, ϵ_0) are the constitutive parameters of free space, and $\hat{n}$ is the unit outward normal to the closed surface S. When radiating in free space, $(\boldsymbol{J}^i, \boldsymbol{M}^i)$ produce the known incident fields denoted $(\boldsymbol{E}^i, \boldsymbol{H}^i)$. If the impressed sources are very close to the conducting body, then the geometry of fig. 2a is referred to as an antenna or radiation problem. If the impressed currents are far removed, then the incident fields are plane waves, and fig. 2a is a scattering problem.

As illustrated in fig. 2b, the first step in obtaining the integral equation is to use Schelkunoff's surface equivalence principle [4,16,17] to replace the perfectly conducting body by free space and the equivalent electric surface current

$$\boldsymbol{J} = \hat{n} \times \boldsymbol{H} \quad \text{on S.} \tag{16}$$

The free space fields of the equivalent current $\boldsymbol{J}$ are referred to as the scattered fields, and are denoted $(\boldsymbol{E}^S, \boldsymbol{H}^S)$. In the equivalent problem of fig. 2b, the total fields are the superposition of the free space fields of $(\boldsymbol{J}^i, \boldsymbol{M}^i)$ and $\boldsymbol{J}$, i.e.

$$\boldsymbol{E} = \boldsymbol{E}^i + \boldsymbol{E}^S, \tag{17}$$

$$\boldsymbol{H} = \boldsymbol{H}^i + \boldsymbol{H}^S. \tag{18}$$

It is important to emphasize that in the equivalent problem of fig. 2b, all currents radiate in free space. As a result, the integral equation and MM solution for $\boldsymbol{J}$ is formulated entirely in terms of the free space fields of currents.

The electric field integral equation (EFIE) is a statement of the boundary condition that the total electric field tangential to S must vanish,

$$-\hat{n} \times \boldsymbol{E}^S = \hat{n} \times \boldsymbol{E}^i \quad \text{on S.} \tag{19}$$

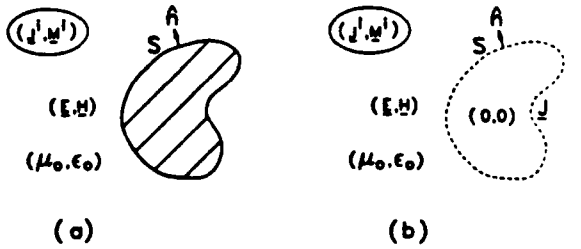


Fig. 2. (a) The impressed currents $(\boldsymbol{J}^i, \boldsymbol{M}^i)$ radiate in the presence of a perfectly conducting body. (b) The body is replaced by the equivalent electric surface current $\boldsymbol{J}$.

Equation (19) is referred to as an integral equation for J, since the scattered field can be written as

$$E^S = \int\int_S J \cdot \bar{G}\, ds, \tag{20}$$

where $\bar{G}$ is the free space electric dyadic Green's function [18].

The magnetic field integral equation (MFIE) is a statement of the surface equivalence theorem (16). Inserting eq. (18) into (16) yields

$$-\hat{n} \times H^S + J = \hat{n} \times H^i \quad \text{on } S^+, \tag{21}$$

where S^+ is a surface an infinitesimal distance exterior to S.

In the above derivations of the EFIE and the MFIE, the use of Schelkunoff's surface equivalence principle requires that S be a *closed surface*. For a closed surface, such as a sphere or a closed box, J is the current flowing on the exterior of S, and there is no current on the interior. However, for an *open surface*, such as a zero thickness plate or a box with one side removed, there will in general be different surface currents flowing on either side. By considering a zero thickness plate as the limiting case of a closed box as the thickness goes to zero, it can be shown that when the EFIE is applied to open surfaces, J will be the vector sum of the current on the two sides [19]. Since it is the vector sum current which radiates the scattered fields, the EFIE is applicable to closed as well as open surfaces. By contrast, when the MFIE is applied to open surfaces the resulting J is the vector difference between the top and bottom currents. The vector difference current has no use except in the special case where the current on one side is zero, i.e. for closed surfaces. Thus, the MFIE is only applicable to closed surfaces.

3.2. The MM solution

This section will describe the MM solution of the basic EFIE of eq. (19). Comparing eq. (19) to (1) it can be seen that:

- the excitation function g is the incident electric field tangential to S;
- the unknown response f is the surface current J flowing on S;
- the linear operator L is minus the tangential to S component of the free space electric field of an electric current.

The first step in the MM solution is to expand the unknown current in terms of some basis functions. Thus we expand J as

$$J \approx J^N = \sum_{n=1}^{N} I_n J_n, \tag{22}$$

where the J_n are a sequence of N known linearly independent vector expansion functions, and the I_n are a sequence of N unknown complex coefficients ($n = 1, 2, \ldots, N$). Note that the J_n must in general account for both components of the vector current J on S. Denoting $-L(J_n) = E_n$ as the free space field of J_n, eq. (19) becomes

$$-\hat{n} \times \sum_{n=1}^{N} I_n E_n = \hat{n} \times E^i \quad (\text{on S}). \tag{23}$$

Now define w_m ($m = 1, 2, \ldots, N$) as a sequence of N linearly independent weighting functions on S, and with vector direction tangential to S. Taking the vector inner product of both sides of eq. (23) with the w_m and integrating over S yields

$$-\sum_{n=1}^{N} I_n \int\int_S E_n \cdot w_m\, ds = \int\int_S E^i \cdot w_m\, ds, \quad m = 1, 2, \ldots, N. \tag{24}$$

Equation (24) is a system of N simultaneous linear equations, which can be more compactly written in matrix form as

$$[Z]I = V, \tag{25}$$

where, in analogy with Ohm's law, $[Z]$ is the order N impedance matrix, V is the length N voltage vector, and I is the length N current or solution vector which contains the I_n in eq. (22). Typical elements of $[Z]$ and V are given by

$$Z_{mn} = -\int\int_m E_n \cdot w_m\, ds, \quad m, n = 1, 2, \ldots, N, \tag{26}$$

$$V_m = \int\int_m E^i \cdot w_m\, ds, \quad m = 1, 2, \ldots, N, \tag{27}$$

where the integration is over that portion of S where w_m is non-zero. Note that the matrix equation (25) requires only the evaluation of the *free space* fields of known currents. If the weighting functions are considered as surface currents with units A/m, then the elements of $[Z]$ and V have units VA, and the I_n are dimensionless. Equation (25) can now be solved for I, which when substituted into eq. (22) provides an approximation to J. Once J is known most parameters of interest, such as radiated or scattered fields, can be easily computed since they are simply the *free space* fields of J.

3.3. TM scattering by a perfectly conducting strip

This section will present the MM solution of the EFIE, eq. (19), for the relatively simple 2D problem of TM scattering by a perfectly conducting strip. In particular, it is desired to obtain simple expressions for the elements in the MM matrix equation that can be coded with a minimum of effort, and evaluated with a minimum of CPU time. Figure 3 shows a TM (to z) polarized plane wave incident upon a perfectly conducting strip of width D. The incident electric field is

$$E^i = \hat{z} e^{jk(x\cos\phi_i + y\sin\phi_i)}, \tag{28}$$

where ϕ_i is the angle of incidence and $k = 2\pi/\lambda$ is the free space wavenumber. For the TM polarization, the electric fields and currents are purely $\hat{z}$ directed. Thus, the vector notation will be dropped and the $\hat{z}$ component is understood. For example, the vector EFIE, eq. (19), reduces to the scalar EFIE

$$-E^s = E^i \quad \text{on S.} \tag{29}$$

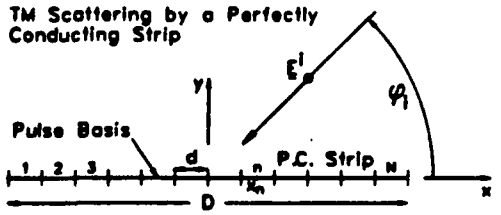


Fig. 3. Geometry for the MM solution of a TM plane wave incident upon a perfectly conducting strip of width D.

Following eq. (22), the current on the strip is expanded as

$$J \approx J^N = \sum_{n=1}^{N} I_n J_n, \tag{30}$$

where as described in section 3.1, J is the sum of the surface current on the top and bottom surfaces of the strip. In order to define the MM basis functions, the strip of width D is split into N smaller strips or segments of width $d = D/N$, and with center point x_n. The expansion functions are chosen as the simple piecewise constant or pulse functions

$$J_n = \frac{1}{d}\text{A/m} \quad \text{on segment } n,$$
$$x_n - \tfrac{1}{2}d \le x \le x_n + \tfrac{1}{2}d,$$
$$= 0, \quad \text{otherwise.} \tag{31}$$

Note that the J_n have been normalized so that they have 1 A total current. For the pulse expansion to be accurate, d must be chosen small enough that the current is essentially constant in each strip. Typically this requires strips of width $0.05\lambda \le d \le 0.25\lambda$, with the accuracy increasing as d/λ decreases.

In order to find Z_{mn}, we will first find E_n. The free space electric field of a unit amplitude line source is

$$E_L = \frac{-\pi\eta}{2\lambda} H_0^{(2)}(k|\rho - \rho'|), \tag{32}$$

where the characteristic impedance of free space, $\eta = \sqrt{\mu_0/\epsilon_0} \approx 377\ \Omega$, and $|\rho - \rho'|$ is the distance from the line source to the field point. Using primed and unprimed coordinates to denote the source and field point, respectively,

$$|\rho - \rho'| = \sqrt{(x-x')^2 + (y-y')^2}$$
$$= \sqrt{\rho^2 + \rho'^2 - 2\rho\rho'\cos(\phi - \phi')}. \tag{33}$$

Using eq. (32) and superposition, the free space field of J_n is

$$E_n(x, y) = \frac{-\pi\eta}{2\lambda} \int_n J_n H_0^{(2)}(k|\rho - \rho'|)\, dx', \tag{34}$$

where the integral is over the width of segment n, i.e. $x_n - \frac{1}{2}d \le x' \le x_n + \frac{1}{2}d$. Employing Galerkin's method, with weighting functions $w_m = J_m$, the elements in the MM matrix equation, eqs. (26) and (27), reduce to

$$Z_{mn} = -\int_m E_n J_m \, dx = \frac{\pi\eta}{2\lambda d^2}\int_m \left[\int_n H_0^{(2)}(k|x-x'|)\, dx'\right] dx, \tag{35}$$

$$V_m = \int_m E^i J_m \, dx = \frac{1}{d}\int_m e^{jkx\cos\phi_i}\, dx = \frac{2\sin(kd\cos\phi_i/2)}{kd\cos\phi_i} e^{jkx_m\cos\phi_i}, \tag{36}$$

where x_m is the center point of segment m, and $m, n = 1, 2, \ldots, N$. Since this is a Galerkin method solution, the impedance matrix is symmetric, i.e. $Z_{mn} = Z_{nm}$.

As is typical for MM solutions, the computation of the N elements of the voltage vector, V, is straightforward and fast (small CPU time). By contrast, the computation of the N^2 elements of the impedance matrix is more complex, and can require a great deal of CPU time. The evaluation of the Z_{mn} involves a double integral which must either be done numerically or by some approximate method. In addition, for self impedance ($m = n$) and adjacent mutual impedance ($|m - n| = 1$) terms, the integrand is singular and can not be evaluated by straightforward numerical integration. Below, we will present simple expressions for the Z_{mn} which can be evaluated without the need for numerical integration.

The self impedance terms are typically the most important and most difficult terms to evaluate in the MM [Z] matrix. Without loss of generality, fig. 4 shows the expansion mode J_n centered about the y axis. Evaluation of the self impedance, Z_{nn}, requires finding the electric field of J_n directly on J_n. The electric field of J_n at a field point ($-d/2 \le x \le d/2$, $y = 0$) on J_n is

$$E_n(x) = \frac{-\pi\eta}{2\lambda}\int_{-d/2}^{d/2} J_n H_0^{(2)}(k|x-x'|)\, dx' = \frac{-\pi\eta}{2\lambda d}\left[\int_{-d/2}^{x} H_0^{(2)}(k(x-x'))\, dx' + \int_x^{d/2} H_0^{(2)}(k(x'-x))\, dx'\right]. \tag{37}$$

Note that the integrands in eq. (37) are singular when $x = x'$, and thus they can not be evaluated by straightforward numerical integration.

Since $d \ll \lambda$, one method for treating the singularity is to replace the Hankel function by its small argument approximation, and then perform the integrals analytically. Using the approximation [20]

$$H_0^{(2)}(u) \approx (1 + j0.0738) - \frac{j2}{\pi}\ln u, \quad |u| \ll 1, \tag{38}$$

in eq. (37) and integrating yields

$$E_n(x) = \frac{-\pi\eta}{2\lambda d}\left[\left(C + j\frac{2}{\pi}\right) d - \frac{j2}{\pi} \times \left((x + \tfrac{1}{2}d)\ln k(x + \tfrac{1}{2}d) + (\tfrac{1}{2}d - x)\ln k(\tfrac{1}{2}d - x)\right)\right], \tag{39}$$

where the complex constant $C = 1 + j0.0738$. Note that $E_n(x)$ in eq. (39) is well behaved and contains no singularities. Finally, inserting (39) into (35) and integrating yields

$$Z_{nn} = -\int_{-d/2}^{d/2} E_n(x) J_n \, dx = \frac{\pi\eta}{2\lambda}\left[\left(C + j\frac{3}{\pi}\right) - \frac{j2}{\pi}\ln kd\right]. \tag{40}$$

Figure 5 shows the self impedance, $Z_{nn} = R_{nn} + jX_{nn}$, versus d/λ computed by eq. (40) and by the

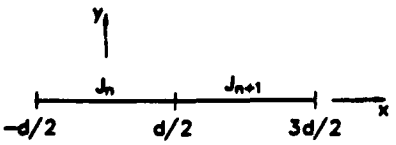


Fig. 4. Geometry for the computation of the self and touching impedances.

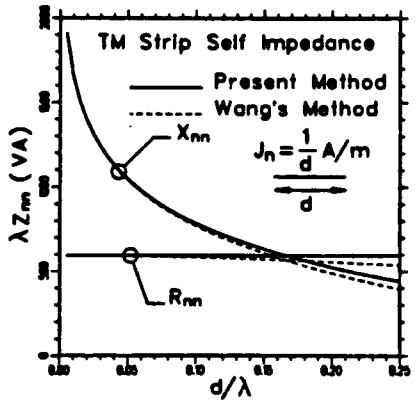


Fig. 5. A comparison of the self impedance, $Z_{nn} = R_{nn} + jX_{nn}$, computed by eq. (40) and by the more exact method of Wang [21].

more exact methods of Wang [21]. Note that eq. (40) is reasonably accurate, even for strips as wide as $kd = \frac{1}{2}\pi$.

Referring to fig. 4, adjacent modes have a common endpoint, and thus their mutual impedance will also involve a singular integrand. Using the same small argument approximation for the Hankel function, the mutual impedance between adjacent modes is

$$Z_{n,n+1} = -\int_{d/2}^{3d/2} E_n(x) J_{n+1}\, dx \simeq \frac{\pi\eta}{2\lambda}\left[\left(C + j\frac{3}{\pi}\right) - \frac{j2}{\pi}(2 \ln 2kd - \ln kd)\right]. \quad (41)$$

If modes m and n do not touch ($|m-n| > 1$) then Z_{mn} can be evaluated using eq. (35) and numerical integration. However, if modes m and n are not too close, then Z_{mn} can be approximated by

$$Z_{mn} = \frac{\pi\eta}{2\lambda} H_0^{(2)} k(|x_m - x_n|), \quad |x_m - x_n| \gg d. \quad (42)$$

A simple MM matrix equation (25) can be set up using eqs. (40), (41) and (42) for the computation of self, touching mutual, and non-touching mutual impedances, respectively, and eq. (36) for the voltage vector. Using standard matrix algebra, eq. (25) can then be solved for the current vector, I, which when substituted into eq. (22) yields J^N which is an approximation to the true current J on the strip. The next section will discuss the computation of the far zone scattered fields and echo width of the strip.

3.3.1. Far zone fields

In scattering problems, normally the parameter of interest is the far zone scattered fields or echo width. The computation of the scattered fields is straightforward since by definition the scattered fields are the *free space* fields of J. Replacing J by its N term approximation J^N,

$$E^S = \sum_{n=1}^{N} I_n E_n, \quad (43)$$

where the coefficients I_n ($n = 1, 2, \ldots, N$) are evaluated by the MM solution described in section 3.3, and E_n is the free space electric field of the basis function, J_n.

This section will derive simple expressions for the scattered fields in the *far zone*, i.e. in the limit as $k\rho \to \infty$. In general, E_n is given by eq. (34). In the far zone, the Hankel function can be replaced by its large argument approximation [20]

$$H_0^{(2)}(u) \approx \sqrt{\frac{2}{\pi u}}\, e^{-j(u-\pi/4)}, \quad |u| \gg 1. \quad (44)$$

In the far zone $\rho \gg \rho'$, and eq. (33) for the distance between source point x' on the strip and the field point (ρ, ϕ) reduces to

$$|\rho - \rho'| \approx \sqrt{\rho^2 - 2\rho\rho' \cos(\phi - \phi')} \approx \rho - x' \cos\phi. \quad (45)$$

Using eq. (45) to approximate the exponential phase term in eq. (44), and $1/|\rho - \rho'| \approx 1/\rho$ for the amplitude, the far zone approximation for the Hankel function becomes

$$H_0^{(2)}(k|\rho - \rho'|) \approx \frac{\sqrt{2}}{\sqrt{\pi k\rho}} e^{j\pi/4}\, e^{-jk(\rho - x'\cos\phi)}, \quad |k\rho| \gg 1 \quad \text{and} \quad \rho \gg \rho'. \quad (46)$$

Inserting eq. (46) into (34) and integrating yields

$$E_n(\rho, \phi) = -\frac{\pi\eta}{2d\lambda}\frac{\sqrt{2}}{\sqrt{\pi k}}\, e^{j\pi/4}\frac{e^{-jk\rho}}{\sqrt{\rho}}\int_n e^{jkx'\cos\phi}\, dx'$$

$$= -\frac{e^{j\pi/4}\eta}{2\sqrt{\lambda}}\frac{e^{-jk\rho}}{\sqrt{\rho}}V_n(\phi). \quad (47)$$

Note that in the far zone E_n is proportional to V_n of eq. (27), and in fact eq. (47) could have been derived from reciprocity rather than by direct integration. Since the ρ dependence of $E_n(\rho, \phi)$ in the far zone is simply

$$\frac{e^{-jk\rho}}{\sqrt{\rho}},$$

$E_n(\rho, \phi)$ is usually written as

$$E_n(\rho, \phi) = \frac{e^{-jk\rho}}{\sqrt{\rho}}E_{nF}(\phi), \quad (48)$$

where $E_{nF}(\phi)$ is the *far zone* electric field of J_n. Comparing eqs. (47) and (48) gives

$$E_{nF}(\phi) = \sqrt{\rho}\, e^{jk\rho}E_n(\rho, \phi) = -\frac{e^{j\pi/4}\eta}{2\sqrt{\lambda}}V_n(\phi). \quad (49)$$

The total far zone scattered electric field is

$$E_F^S(\phi) = \sum_{n=1}^{N} I_n E_{nF}(\phi) = -\frac{e^{j\pi/4}\eta}{2\sqrt{\lambda}}\sum_{n=1}^{N} I_n V_n(\phi). \quad (50)$$

For 2D problems, the power density of the far zone scattered field is usually expressed in terms of the echo width. The echo width, W, is defined as a width, when multiplied by the power density of the incident wave, that would yield sufficient power that could produce by isotropic radiation the same radiation intensity as that in a given direction from the scattering object. For the unit amplitude incident plane wave being considered here, the echo width is related to the far zone field by

$$W(\phi) = 2\pi\left|E_F^S(\phi)\right|^2. \quad (51)$$

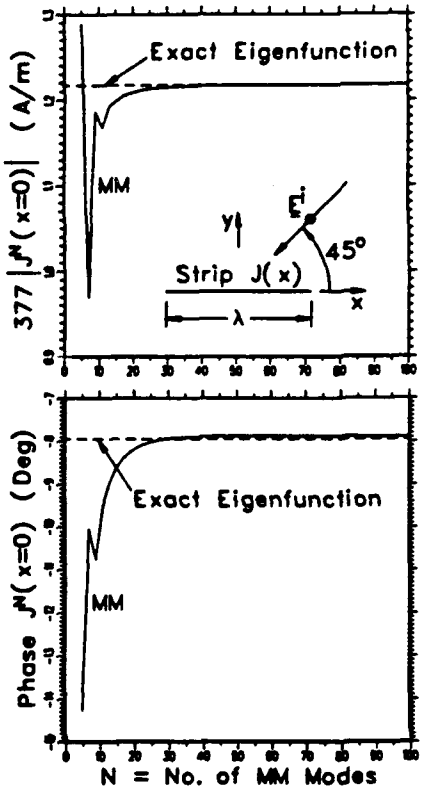


Fig. 6. A convergence curve for the current in the center of the strip.

W has dimensions of meters, and is often expressed in terms of dB m which is obtained by taking $10 \log_{10} W$.

3.3.2. Numerical results

This section will present numerical results based upon the above MM solution for TM scattering by a perfectly conducting strip. This data will include strip current distribution, echo width, and CPU times, and will be designed to illustrate the accuracy and convergence of the MM solution. As illustrated in the insert in fig. 6, all data in this section will be for a 1 λ wide strip illuminated by a TM wave incident from the angle $\phi_i = 45°$ with respect to the x axis.

The MM solution basically determines $J^N(x)$, an N term approximation to the current induced on the strip by the incident field. Figure 6 shows the magnitude and phase of the current in the center of the strip, $J^N(x = 0)$, versus N = the

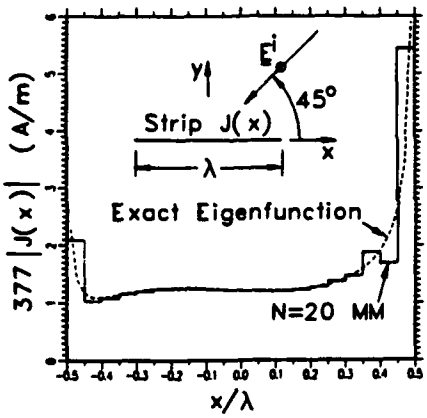


Fig. 7. A comparison of the $N = 20$ MM and eigenfunction solution for the current on the strip.

number of pulse basis functions used in eq. (30) to expand the strip current. The dashed line shows the exact eigenfunction solution for the strip current [22,23]. For small N, the MM solution is erratic, however, as N increases it converges almost exactly to the eigenfunction solution. For $N = 20$, which corresponds to a segment size of 0.05λ, fig. 7 shows a comparison of the magnitude of strip current computed by the MM and the eigenfunction solution. The agreement is very good near the center of the strip, however, it does worsen near the edges where the true current has a $1/\sqrt{s}$ singularity (s = distance to the strip edge).

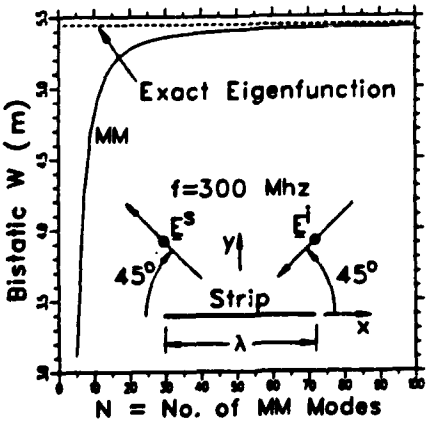


Fig. 8. A convergence curve for the bistatic echo width of the strip.

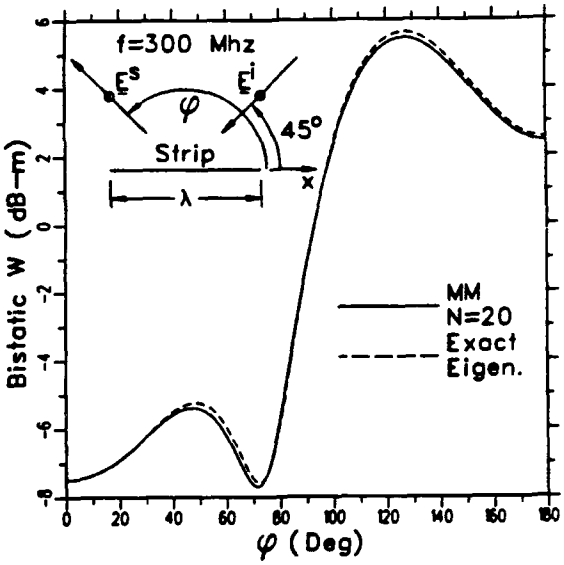


Fig. 9. A comparison of the $N = 20$ MM and eigenfunction solution for the bistatic scattering from the strip.

In most cases one is more interested in the echo width than the strip current. Figure 8 shows the bistatic echo width at $\phi = 135°$ versus N. Note that it converges uniformly to the eigenfunction solution shown as the dashed line. Figure 9 shows a comparison of the eigenfunction and $N = 20$ mode MM solution for the bistatic scattering pattern of the strip. The agreement is less than a few tenths of dB for all angles.

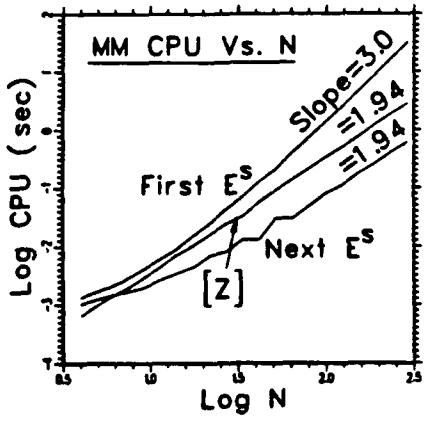


Fig. 10. The CPU time on a VAX 8550 for various parts of the MM solution.

The final set of data will illustrate the CPU time for the strip MM solution. As a function of N, fig. 10 shows the CPU time to:

1. compute the $N \times N$ impedance matrix, $[Z]$;
2. make the first computation of the scattered field, E^S, which involves (1) computing the voltage vector, V, using eq. (36), (2) LU decomposing the $[Z]$ matrix, (3) back substituting to find the solution vector I, and (4) finding the far zone scattered fields using eq. (50);
3. make a second or subsequent computation of E^S which requires the above 4 steps except step (2), the LU decomposition of $[Z]$.

For large N, the CPU time is dominated by the first computation of the scattered field. Of the four steps listed in item 2 above, the LU decomposition of the $[Z]$ matrix requires by far the largest CPU time. For large N the "first E^S" curve is nearly a straight line with slope ~ 3.0, indicating that the LU decomposition is an N^3 process. The next largest CPU time is that to compute the impedance matrix $[Z]$. Since there are N^2 elements in $[Z]$ this is an N^2 process. For large N, the slope of this curve is about 1.94. It should be noted that the expressions used here to evaluate the $[Z]$ matrix are extremely simple and fast to evaluate. In a more typical situation the computation of the elements of $[Z]$ require one or more integrals which must be evaluated numerically. In this case, the CPU time would be dominated by the computation of the $[Z]$ matrix for small N, and the LU decomposition for large N. The smallest CPU time is that for a second or subsequent computation of the scattered field, since the $[Z]$ matrix has already been LU decomposed. However, the total CPU time to compute a pattern can be significant, since E^S must be evaluated at many angles. For example, computation of a backscatter pattern at 1° steps for $N = 300$, will require 3 s to compute the $[Z]$ matrix, 33 s for the first computation of E^S, and 360×0.63 s = 227 s for the remaining 360 angles.

4. Material bodies

This section will outline the volume integral equation and MM solution to the problem of scattering by a dielectric and/or ferrite material body. The method will be illustrated by the relatively simple problem of TM scattering by a rectangular dielectric cylinder.

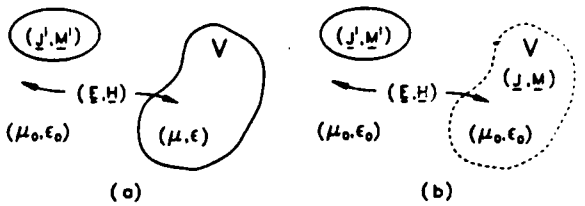


Fig. 11. (a) The sources (J^i, M^i) radiate the fields (E, H) in the presence of a material body; (b) the material body is replaced by free space and the equivalent current (J, M).

4.1. The volume integral equation

This section will obtain the volume integral equation for scattering by a material body. The general problem is illustrated in fig. 11a where the assumed known impressed currents (J^i, M^i) radiate the known incident fields (E^i, H^i) in free space, and the unknown total fields (E, H) in a medium which is free space except for the presence of a possibly lossy and inhomogeneous dielectric/ferrite material body with constitutive parameters (μ, ϵ) in the volume V.

As illustrated in fig. 11b, the volume equivalence theorem is used to replace the material body by free space and by the equivalent electric and magnetic volume polarization currents [4]

$$J = j\omega(\epsilon - \epsilon_0)E \quad \text{in V}, \tag{52}$$

$$M = j\omega(\mu - \mu_0)H \quad \text{in V}. \tag{53}$$

Note that J and M exist only in the volume V where $\epsilon \neq \epsilon_0$ and $\mu \neq \mu_0$, respectively. In the equivalent problem of fig. 11b the total fields at any point in space (interior or exterior to V) are the superposition of the free space fields of the impressed currents and the equivalent currents, i.e.

$$E = E^i + E^J + E^M, \tag{54}$$

$$H = H^i + H^J + H^M, \tag{55}$$

where (E^J, H^J) and (E^M, H^M) are the free space fields of J and M, respectively.

The volume integral equation is obtained by enforcing the volume equivalence theorems in V. Solving eqs. (52) and (53) for E and H, and substituting the results into eqs. (54) and (55) yields

$$-E^J + \frac{J}{j\omega(\epsilon - \epsilon_0)} - E^M = E^i \quad \text{in V}, \tag{56}$$

$$-H^J + \frac{M}{j\omega(\mu - \mu_0)} - H^M = H^i \quad \text{in V}. \tag{57}$$

Equations (56) and (57) are a pair of coupled vector integral equations for (J, M). For nonmagnetic material ($\mu = \mu_0$ and thus $M = 0$), eq. (56) is equivalent to the three scalar equations

$$-E_x^{J_x} - E_x^{J_y} - E_x^{J_z} + \frac{J_x}{j\omega(\epsilon - \epsilon_0)} = E_x^i, \tag{58}$$

$$-E_y^{J_x} - E_y^{J_y} - E_y^{J_z} + \frac{J_y}{j\omega(\epsilon - \epsilon_0)} = E_y^i,$$

$$-E_z^{J_x} - E_z^{J_y} - E_z^{J_z} + \frac{J_z}{j\omega(\epsilon - \epsilon_0)} = E_z^i,$$

for (J_x, J_y, J_z) in V. Figure 12 shows a rectangular dielectric volume which has been segmented into a number of smaller rectangular cells for the purpose of defining the MM expansion functions. Each cell contains three MM basis functions, corresponding to $\hat{x}$, $\hat{y}$, and $\hat{z}$ components of J. Thus, M cells will result in $N = 3M$ unknowns. Employing the MM with three orthogonal vector weighting functions in each cell will reduce eq. (58) to an order $3M$ matrix equation.

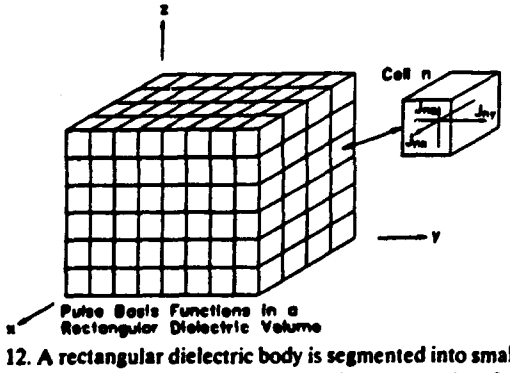


Fig. 12. A rectangular dielectric body is segmented into smaller rectangular cells corresponding to the MM expansion functions.

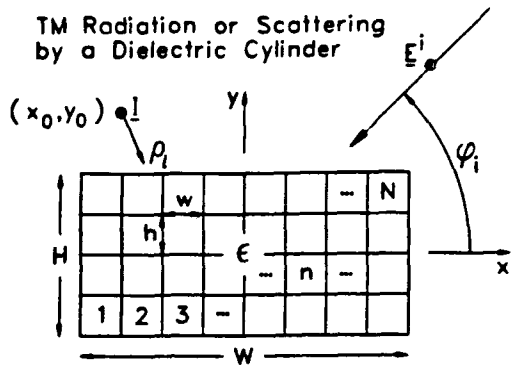


Fig. 13. A rectangular dielectric cylinder is segmented into N rectangular cells corresponding to the MM expansion modes.

4.2. TM dielectric cylinder

This section will outline the MM solution for TM radiation or scattering by a rectangular dielectric cylinder. Figure 13 shows a rectangular cylinder of height H, width W, and permittivity ϵ. A lossy dielectric is described by its complex permittivity

$$\epsilon = \epsilon_r\epsilon_0(1 - j\tan\delta) = \epsilon_r\epsilon_0 - j\frac{\sigma}{\omega}, \tag{59}$$

where ϵ_r is the relative permittivity, $\tan\delta$ is the loss tangent, and σ is the conductivity. For inhomogeneous dielectrics ϵ will be a function of position. Tne dielectric is shown excited by either a TM polarized plane wave with incident electric field given by eq. (28), or by a unit amplitude ($I = \hat{z}$ A) line source located at (x_0, y_0) and with incident electric field

$$E^i = \hat{z}\frac{-\pi\eta}{2\lambda}H_0^{(2)}(k\rho_i), \tag{60}$$

where the local ρ_i is the distance from the line source to the field point.

For a TM to z polarized incident electric field all electric fields and electric currents will be $\hat{z}$ directed. Thus, we can now drop the vector notation, and the $\hat{z}$ component is understood. The

general vector integral eq. (56) reduces to the scalar equation

$$-E^J + \frac{J}{j\omega(\epsilon - \epsilon_0)} = E^i \quad \text{in R}, \tag{61}$$

for the $\hat{z}$ component of J. Since this is a 2D problem, eq. (61) applies in the cross-sectional region R of the cylinder.

The first step in the MM solution of eq. (61) is to expand the unknown volume polarization current as

$$J \approx J^N = \sum_{n=1}^{N} I_n J_n, \tag{62}$$

where the J_n are a sequence of N known linearly independent expansion functions, and the I_n are a sequence of N unknown complex coefficients ($n = 1, 2, \ldots, N$). To define the expansion functions, the rectangular cylinder is segmented into N smaller rectangular cells. If we denote R_n as the cross-section region of cell n, then the piecewise constant or pulse expansion functions are defined by

$$J_n = \frac{1}{A} \text{ A/m}^2 \quad \text{in } R_n,$$
$$= 0 \quad \text{otherwise}, \tag{63}$$

where A is the cross-section area of the cells. As was the case for the strip modes defined in eq. (31), the dielectric expansion modes are normalized to have unit total current.

The free space electric field of J_n at the field point (x, y) is

$$E_n(x, y) = \frac{-\pi\eta}{2\lambda} \int\int_n J_n H_0^{(2)}(k\rho)\, ds', \tag{64}$$

where $\rho = \sqrt{(x-x')^2 + (y-y')^2}$ is the distance from the source point to the field point and the double integral is over R_n, the region of cell n. Using this notation, eq. (61) becomes

$$-\sum_{n=1}^{N} I_n E_n + \sum_{n=1}^{N} \frac{I_n J_n}{j\omega(\epsilon - \epsilon_0)} = E^i \quad \text{in R}. \tag{65}$$

The weighting functions will be Dirac delta functions located at the center of the N cells,

$$w_m = \delta(x - x_m)\, \delta(y - y_m), \quad m = 1, 2, \ldots, N, \tag{66}$$

where (x_m, y_m) is the center of cell m. Multiplying both sides of eq. (65) by the w_m ($m = 1, 2, \ldots, N$) and integrating over R will reduce eq. (65) into an $N \times N$ system of simultaneous linear equations which can be compactly written in matrix form as

$$[Z + \Delta Z]I = V. \tag{67}$$

Using the sampling property of the Delta function, the elements of the MM matrix equation are simply

$$Z_{mn} = -E_n(x_m, y_m), \tag{68}$$

$$\Delta Z_{mn} = \frac{1}{jA\omega(\epsilon_n - \epsilon_0)}, \quad m = n,$$
$$= 0, \quad m \neq n, \tag{69}$$

$$V_m = E^i(x_m, y_m) \quad \text{for general excitation}, \tag{70}$$
$$= e^{jk(x_m \cos\phi_i + y_m \sin\phi_i)}$$
for plane wave excitation, (71)

$$= \frac{-\pi\eta}{2\lambda} H_0^{(2)}\left(k\sqrt{(x_0 - x_m)^2 + (y_0 - y_m)^2}\right)$$
for line source excitation, (72)

for $m, n = 1, 2, \ldots, N$. Note that the only place that the permittivity enters the MM solution is in the evaluation of the diagonal $[\Delta Z]$ matrix. Inhomogeneous dielectrics are treated by simply using ϵ_n, the value of ϵ at the center of cell n, in evaluating the ΔZ_{nn}. Also note that the difference between plane wave and line source excitation is a simple change in the voltage vector, V.

Choosing the weighting functions as Dirac delta functions is equivalent to enforcing the integral equation at N points at the center of the N cells, and is therefore referred to as a *point matching* MM solution. The advantage of the point matching solution is that it simplifies the evaluation of the MM matrix equation. For example, evaluating the Z_{mn} for point matching requires a double integral to find $E_n(x_m, y_m)$. By

contrast, in a Galerkin MM solution ($w_m = J_m$) a quadruple integration would be required to find the Z_{mn}. The disadvantage of the point matching solution is that, as compared to Galerkin's method, it tends to be more slowly converging.

The double integral required to find E_n or Z_{mn} must be done either numerically or by some approximate technique. For off diagonal terms ($m \neq n$), numerical integration is reasonably fast and accurate. In fact, when the separation between the cells is much larger than the cell size, J_n can be replaced by a unit amplitude line source at the center of cell n, and the off diagonal terms can be approximated by

$$Z_{mn} \approx \frac{\pi\eta}{2\lambda} H_0^{(2)}(k\rho_{mn}), \tag{73}$$

where $\rho_{mn} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}$ is the distance between the centers of cells m and n. However, for the self impedance terms ($m = n$) the integrand is singular and straightforward numerical integration fails. Providing that the cells are nearly square, Richmond presented a very simple method for evaluating the self impedance integrations [24]. Basically, Richmond used the approximation that the self impedance of a nearly square cell is the same as that of a circular cell of the same cross section area. For a circular cell of radius $a = \sqrt{A/\pi}$ the point matching self impedance is

$$Z_{nn} = -E_n(\rho = 0) = \frac{\pi\eta}{2\lambda} \int\int_n J_n H_0^{(2)}(k\rho)\, ds'$$

$$\approx \frac{\pi\eta}{2A\lambda} \int_0^{2\pi} \int_0^a H_0^{(2)}(k\rho)\rho\, d\rho\, d\phi. \tag{74}$$

In eq. (74), the $d\phi$ integration simply results in a factor of 2π, and the $d\rho$ integration can be integrated using the identity [25]

$$\int \rho H_0^{(2)}(k\rho)\, d\rho = \frac{\rho}{k} H_1^{(2)}(k\rho). \tag{75}$$

In this case, eq. (74) becomes

$$Z_{nn} = \frac{\eta}{2a^2}\left[aH_1^{(2)}(ka) - j\frac{2}{\pi k}\right]. \tag{76}$$

4.3. Numerical results

This section will present numerical results based upon the point matching MM solution for scattering by a rectangular dielectric cylinder given in section 4.2. For simplicity, self impedance terms are evaluated using eq. (76) and all other mutual impedances using eq. (73).

Figure 14 illustrates the convergence of the backscatter echo width, W, at $f = 300$ Mhz for a 1×0.5 m dielectric cylinder with relative permittivity $\epsilon_r = 4$ and loss tangent $\tan\delta = 0.1$. Curves are shown for the dielectric cylinder segmented into $N = 6 \times 3 = 18$, $10 \times 5 = 50$, $20 \times 10 = 200$, and $30 \times 15 = 450$ cells. The CPU times for these runs were approximately 1.63, 5.27, 55.79, and 319.15 s, respectively, on a VAX 8550. The MM solution is well converged for the 20×10 segmentation, which corresponds to a segment size of 0.05λ or 0.1 wavelength in the dielectric.

One of the useful features of the volume current formulation is that it is trivial to treat inhomogeneous dielectrics and to compute the total electric field in the dielectric. Inhomogeneous dielectrics are treated by simply using ϵ_n = the value of ϵ at the center of cell n in evaluating the

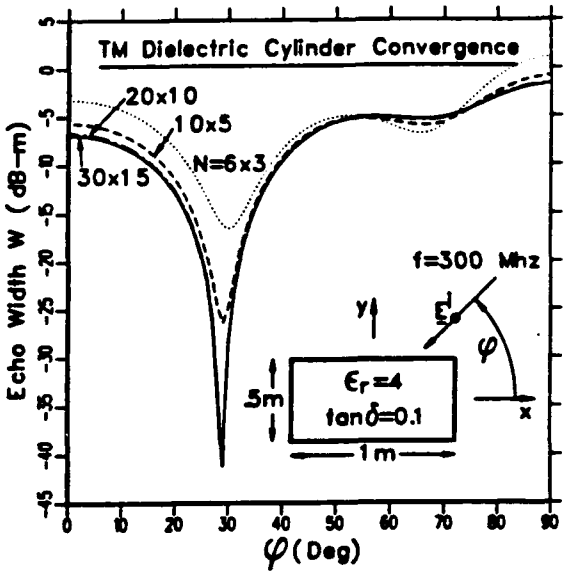


Fig. 14. Convergence of the MM solution for TM scattering by a rectangular dielectric cylinder.

ΔZ_{nn} terms in eq. (69). Using eqs. (52) and (62), the total electric field in the center of cell n is simply

$$E(x_n, y_n) = \frac{I_n}{j\omega(\epsilon_n - \epsilon_0)A}, \tag{77}$$

where I_n is the coefficient of J_n and A is the cross-section area of the cells.

The insert in fig. 15 shows an $f = 1$ Ghz TM plane wave with edge on incidence ($\phi_i = 0$) to a thin dielectric slab of width 0.6 m and thickness 0.0075 m. For the MM solution, the slab is segmented into a single row of $N = 80$ cells. The magnitude of the total internal fields along the slab centerline is shown for a homogeneous dielectric slab with $\epsilon_r = 16$, and for an inhomogeneous slab with

$$\epsilon_r = 1 + \frac{15}{0.3}(0.3 - |x|), \tag{78}$$

in which ϵ_r tapers linearly from 16 in the center of the slab to 1 at the edges. When the incident waves hits the edge of the homogeneous $\epsilon_r = 16$ slab, a surface wave is produced by the abrupt change in permittivity seen by the incident field. The surface wave propagates toward the other edge of the slab and is then largely reflected. The result is the oscillatory or standing wave pattern

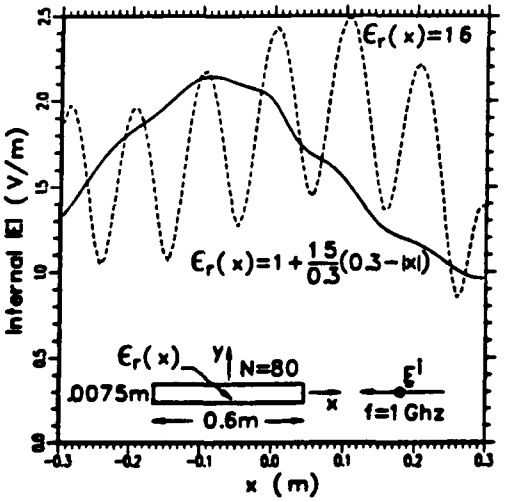


Fig. 15. The total internal fields in a homogeneous (dashed line) and inhomogeneous (solid line) dielectric slab.

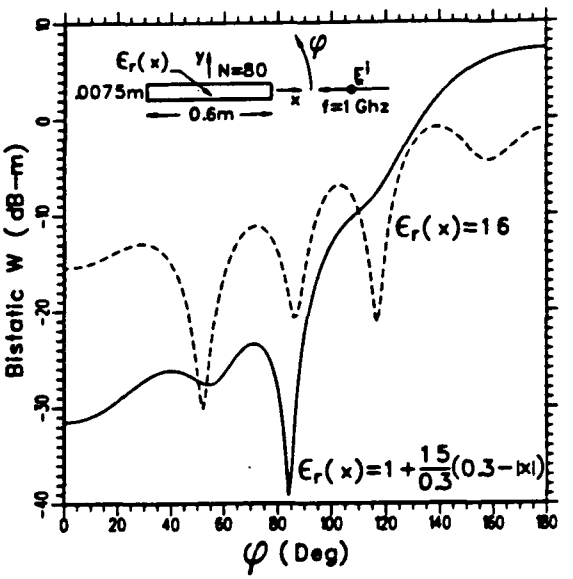


Fig. 16. The bistatic echo width for a homogeneous (dashed line) and inhomogeneous (solid line) dielectric slab.

shown for the internal fields. By contrast, since the edges of the inhomogeneous slab have $\epsilon_r = 1$, very little surface wave is produced, and the internal fields display almost no oscillatory behavior. The bistatic echo widths for the homogeneous and inhomogeneous slabs are shown in fig. 16.

5. The internal resonance problem

Every technique, no matter how powerful and general, has its limitations and problems. This section will describe the internal resonance problem which occurs in MM solutions for currents on closed surfaces.

In solving eq. (1) for f it is implicit that we desire the *particular* or *forced response*, i.e. the response due to the excitation g. However, if there is a solution to the homogeneous equation $L(f) = 0$, then the total response is $f = f_p + cf_0$ where f_p is the desired particular solution, f_0 is the homogeneous solution, and c is an arbitrary constant. If the operator equation has a homogeneous solution, then ideally, the MM matrix eq.

(7) will also have a homogeneous solution. The rank of the matrix $[L]$ will be less than N, and $[L]$ will be a singular matrix with a zero determinant and no inverse. However, since the matrix $[L]$ is a finite-dimensional approximation to the operator L, in practice, $[L]$ will be nearly singular with a small but non-zero determinant and a very large condition number. In this case, the numerical solution of eq. (7) will be extremely sensitive to roundoff error, and obtaining an accurate solution will be difficult if not impossible. Thus, if the original operator equation has a homogeneous solution, the MM solution is expected to fail.

In electromagnetics, homogeneous solutions arise in the analysis of closed surfaces. For example, consider the problem of scattering by a closed perfectly conducting box. The current on the box is the solution of the EFIE, eq. (19), which basically enforces the boundary condition that the total electric field tangential to the surface of the box must be zero. However, it is well known that a cavity mode can exist in the interior of a closed perfectly conducting surface [26]. A cavity mode is a solution of Maxwell's source free equations, which satisfy all boundary conditions. Thus, a cavity mode is a solution of the homogeneous EFIE. If the operating frequency is close to the *internal resonance* frequency of one of the cavity modes of the box, the result will be an ill conditioned MM matrix equation and a failure of the MM solution. The failure can usually be recognized by the erratic behavior of the MM solution over a narrow frequency range [27].

Mautz and Harrington found that the internal resonance problem is a result of the fact that, at frequencies corresponding to internal cavity resonances, the exact solutions to the EFIE and the MFIE are not unique, and thus their MM solutions fail [28]. However, they showed that the *combined-field integral equation* (CFIE), which is obtained via a linear combination of the EFIE and the MFIE, does have a unique solution at all frequencies. As a result, the MM solution of the CFIE does not suffer from the internal resonance problem. Note that the internal resonance problem is more a problem with the non-uniqueness of the operator equation, than it is problem with the MM solution [29].

6. Summary

This paper has presented an introduction to the MM and its application to problems of electromagnetic radiation and scattering. The MM is a numerical technique which is used to solve the linear integral equations which arise in electromagnetics by transforming them into a system of simultaneous linear algebraic equations, i.e. a matrix equation. The method is illustrated by considering both perfectly conducting and dielectric bodies. In particular simple MM solutions were presented for TM scattering by a perfectly conducting strip and a rectangular dielectric cylinder. Although there is no mathematical guarantee of convergence, as is typical, the solutions presented did convergence.

References

[1] R.F. Harrington, Field Computations by Moment Methods (Macmillan, New York, 1968).
[2] R.F. Harrington, Matrix methods for field problems, Proc. IEEE 55 (1967) 136.
[3] R.C. Hansen, Moment Methods in Antennas and Scattering (Artech, Boston, 1990).
[4] C.A. Balanis, Advanced Engineering Electromagnetics (Wiley, New York, 1989) ch. 7.
[5] E.H. Newman, A user's manual for the electromagnetic surface patch code, Ohio State University, Dept. of Electrical Engineering, ElectroScience Lab Report 716199-11, prepared under Grant NSG 1498 with the National Aeronautics and Space Administration, Hampton, VA (August 1988).
[6] E.H. Newman, Polygonal plate modeling, Electromagnetics 10 (1990) 65.
[7] E.H. Newman, P. Alexandroupoulos and E.K. Walton, Polygonal plate modeling of realistic structures, IEEE Trans. Antennas Propag. 32 (1984) 742.
[8] G.J. Burke and A.J. Poggio, Numerical electromagnetic code (NEC) – method of moments, Naval Ocean Systems Center, San Diego, CA, Tech. Doc. 16 (July 1977).
[9] S.M. Rao, D.R. Wilton and A.W. Glisson, Electromagnetic scattering by surfaces of arbitrary shape, IEEE Trans. Antennas Propagat. 30 (1982) 410.
[10] W.A. Johnson and D.R. Wilton, Modeling scattering from and radiation by arbitrary shaped objects with the electric field integral equation triangular surface patch code, Electromagnetics 10 (1990) 41.
[11] E.K. Miller, A selective survey of computational electromagnetics, IEEE Trans. Antennas Propag. 36 (1988) 1281.

[12] T.K. Sarkar, A note on the variational method (Raleigh–Ritz), Galerkins method, and the method of least squares, Radio Sci. 18 (1983) 1207.
[13] D.G. Dudley, Error minimization and convergence in numerical methods, Electromagnetics 5 (1985) 89.
[14] I. Statgold, Boundary Value Problems of Mathematical Physics, Vol. II (MacMillan, London, 1968).
[15] H. Shamansky and C. Yang, Test functions for the moment method which yield the minimum mean square error, Radio Sci. 26 (1991) 133.
[16] R.F. Harrington, Time-Harmonic Electromagnetic Fields (McGraw-Hill, New York, 1961) section 3–7.
[17] S.A. Schelkunoff, On diffraction and radiation of electromagnetic waves, Phys. Rev. 56 (1939).
[18] C.T. Tai, Dyadic Green's Functions in Electromagnetics (Intext, Scranton, PA, 1971).
[19] E.H. Newman and M.R. Schrote, On the current distribution for open surfaces, IEEE Trans. Antennas Propag. 31 (1983) 515.
[20] M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, Applied Mathematics Series 55 (National Bureau of Standards, Washington, DC, 1970) ch. 9.
[21] N.N. Wang, Reaction formulation for radiation and scattering from plates, corner reflectors, and dielectric-coated cylinders, Ph.D. dissertation, The Ohio State University, Dept. of Electrical Engineering (1974).
[22] J.S. Asvestas and R.E. Kleinman, Electromagnetic and Acoustic Scattering by Simple Shapes, J.J. Bowman, T.B.A. Senior and P.L.E. Uslenghi, eds. (Hemisphere, New York, 1987) ch. 4.
[23] A.K. Dominek, A computer code to generate the strip eigenfunction data, The Ohio State University Electro-Science Lab., private communication.
[24] J.H. Richmond, Scattering by a dielectric cylinder of arbitrary cross section shape, IEEE Trans. Antennas Propag. 13 (1965) 334.
[25] I.S. Gradshteyn and I.M. Ryzhik, Tables of Integrals, Series and Products (Academic, New York, 1980).
[26] J. Van Bladel, Electromagnetic Fields (Hemisphere, New York, 1985) ch. 10.
[27] C. Klein and R. Mittra, Stability of matrix equations arising in electromagnetics, IEEE Trans. Antennas Propag., 21 (1973) 902.
[28] J.R. Mautz and R.F. Harrington, *H*-field, *E*-field, and combined-field solutions for conducting bodies of revolution, Arch. Elek. Ubertragung 32 (1978) 159.
[29] E. Arvas and J.R. Mautz, On the non-uniqueness of the surface EFIE applied to multiple conducting and/or dielectric bodies, Arch. Elek. Ubertragung 42 (1988) 364.

Computer Physics Communications 68 (1991) 353–365
North-Holland

Computer Physics Communications

Two ray shooting methods for computing the EM scattering by large open-ended cavities *

R.J. Burkholder, R.-C. Chou and P.H. Pathak
The Ohio State University ElectroScience Laboratory, 1320 Kinnear Road, Columbus, OH 43212, USA

Received 14 January 1991, in revised form 15 May 1991

Two ray shooting approaches are presented for analyzing the high frequency external EM scattering by open-ended waveguide cavities of relatively arbitrary shape and with a planar interior termination. The contribution to the external scattering arising from the interior cavity region is found by (i) the shooting and bouncing ray (SBR) method, and (ii) the generalized ray expansion (GRE) method. The basic difference between the two methods is in the way the rays are initially launched into the cavity interior via the open end which is directly illuminated. While the SBR method tracks only the incident geometrical optics field which enters the cavity, the GRE method also intrinsically includes the fields diffracted into the cavity by the edges at the open end. Also, in the SBR method a new set of rays needs to be tracked for each incidence angle, while in the GRE method only one set of rays needs to be tracked independent of the incidence angles, although this is generally a much larger set than that used in SBR for a *single* incidence angle. Therefore, it is found that the SBR method is preferable for analyzing the scattering from cavities which are very large electrically and for relatively few incidence angles. On the other hand the GRE method is preferable for analyzing not only large but also moderately large cavities and for cases where a large number of incidence angles are needed. Numerical results and comparisons based on these methods are presented.

1. Introduction

This paper presents two ray-based approaches for analyzing the complex problem of high frequency electromagnetic (EM) scattering by relatively arbitrarily shaped perfectly conducting open-ended waveguide cavities with a planar interior termination as illustrated in fig. 1. The two methods are: (i) the shooting and bouncing ray (SBR) method [1–3], and (ii) the generalized ray expansion (GRE) method [3,4]. The excitation is assumed to be an external plane wave, and the medium surrounding the cavity is assumed to be free space. It is also assumed that the geometrical properties of the waveguide cavity are slowly varying and that the interior walls of the cavity may have a thin material coating. While the existence of a material coating can support surface waves excited by the diffraction at the aperture edge, the present ray tracing methods do not take this into account. It is assumed in the present work that the material coating is sufficiently lossy so that the surface waves are attenuated and their effects are therefore negligible. An $e^{j\omega t}$ time dependence is assumed and suppressed in the analysis to follow.

Referring to fig. 1, the total external scattered field can be expressed within the high frequency approximation as

$$\boldsymbol{E}^s = \boldsymbol{E}^s_{rim} + \boldsymbol{E}^s_{cav} + \boldsymbol{E}^s_{ext}, \tag{1}$$

where $\boldsymbol{E}^s_{rim}$ is the field scattered by only the rim edge of the aperture at the open end of the cavity, and $\boldsymbol{E}^s_{cav}$ is the contribution to the scattering from the interior of the cavity. The remaining contribution $\boldsymbol{E}^s_{ext}$ arising from the other exterior

* The work in this paper was sponsored by the NASA/Lewis Research Center (Grant NAG3-476), Joint Services Electronics Program (Contract N00014-89-J-1007), NASA-Ames Research Center (Contract NCA 2-322), and the Ohio Supercomputer Center.

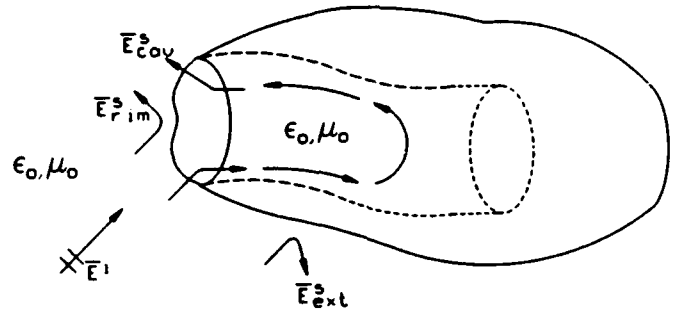


Fig. 1. Open-ended cavity geometry and the associated high frequency scattering mechanisms.

features of the cavity is not of interest since it would depend on the housing in which the cavity is placed, and will therefore not be considered in the present work; in many instances it can be found from the geometrical theory of diffraction (GTD) [5] and its uniform extension (UTD) [6,7] or by the physical theory of diffraction (PTD) [8]. One notes that in eq. (1), E^s_{cav} is usually more dominant than E^s_{rim} for electrically large cavities; this is because the interior wave reflections from the termination are generally significantly larger than the interior waves reflected from the open end. Therefore, the effects of multiple wave interactions between the open end and the termination can be ignored. Likewise, for an electrically large aperture at the open end, the multiple wave diffractions across the aperture contribute weakly to E^s_{rim} and are also ignored. It is further assumed that the angles of incidence and scattering are limited essentially to a half cone angle of about 70° or so about the waveguide axis at the open end. Outside of this region the scattering by the external features of the cavity would dominate. The E^s_{rim} is found via the equivalent current method (ECM) [7] combined with the GTD as discussed in the appendix to this paper. This method of finding E^s_{rim} is presented in more detail in ref. [2]. The scattering mechanism of interest in this paper is that due to the cavity interior E^s_{cav}. After a background discussion of this area, two ray shooting methods for finding E^s_{cav}, SBR and GRE, will be described in detail.

Previously, the E^s_{cav} has been found via hybrid modal, ray and beam techniques as discussed in refs. [1–3]. Cavities which can be adequately modeled by joining together piecewise separable waveguide sections can be analyzed via a hybrid combination of asymptotic high frequency and modal techniques [2,3,9,10]. The modal junction reflection and transmission coefficients, which are associated with the discontinuities formed by joining together the different waveguide sections, are found efficiently via high frequency techniques such as the GTD and the PTD rather than via the classical mode matching technique, which is numerically less efficient for very large apertures. The hybrid modal method gives highly accurate results for a wide variety of waveguide cavity configurations and is therefore often used as a reference solution to test the more approximate but versatile ray and beam based methods. However, as with any modal approach, it becomes cumbersome for guides with very large cross-sections (where a large number of modes are present) and/or material coated interior walls.

In the case of more arbitrarily shaped cavities for which modes cannot even be defined in the conventional sense, the geometrical optics (GO) ray shooting approach proves to be highly useful [1–3]. This approach is frequently referred to as the SBR method and will be described in detail in section 2 of this paper. In this approach, the GO part of the incident plane wave captured by the aperture at the open end is divided into a dense grid of parallel ray tubes, as illustrated in fig. 2. These ray tubes are tracked within the cavity through multiple reflections from the interior cavity walls and planar termination using the laws of GO. The ray tubes eventually exit through

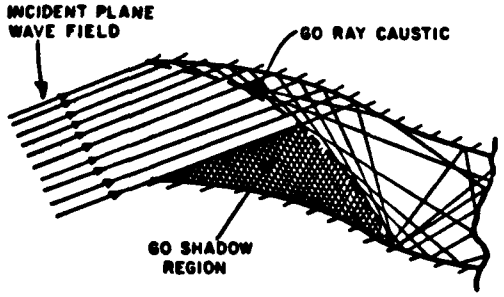


Fig. 2. GO ray tubes launched into an open-ended cavity as in the SBR method which demonstrate a ray caustic and a shadow region.

the aperture where they define equivalent (Huygen's) sources which radiate the fields scattered from the interior of the cavity E^s_{cav} as can be found via aperture integration (AI). However, this GO/AI (or SBR) approach yields a cruder approximation than the hybrid modal method because GO neglects the fields coupled into the interior via diffraction by the edges at the open end, and because it remains reasonably accurate only as long as other interior diffraction effects are weaker than the GO effects. Nevertheless, this approach predicts the dominant high frequency scattering behaviour and can also generally predict the peak envelope of the backscattered field quite well.

To overcome some of the limitations of the SBR method for analyzing arbitrary cavities, a Gaussian beam (GB) shooting technique can be used [2,3]. In this method, the fields incident at the open end of the cavity are expanded in terms of well focussed, shifted, tilted GBs. Once launched, these beams are then tracked inside the slowly varying waveguide cavity only along their beam axes, similar to GO rays. However, one needs far fewer beams than the number of rays which would need to be tracked in a GO ray shooting approach, and the beams overcome the failure of GO at ray caustics. Furthermore, a new set of rays needs to be tracked for each incidence angle in the SBR method, whereas the GBs need to be tracked only once independent of the incidence angles and the GB expansion intrinsically includes the fields coupled to the interior via diffraction of the incident wave by the open end. The main limitation of the axial tracking approximation used in this GB method is that it cannot adequately account for beam distortion the farther the beams propagate inside a curved cavity, thus limiting the method to relatively shallow cavity geometries. Alternatively, the GBs can be tracked more rigorously via complex rays [11], without the limitations of the axial beam tracking approximation. However, this method is far more cumbersome than the axial tracking method which essentially tracks the beam axes like real rays.

More recently, the GRE method (described in detail in section 3 of this paper) has been developed to retain many of the useful features of the Gaussian beam approach and to overcome the problems of beam distortion due to successive reflections in the latter approach [3]. Of course, this GRE approach is not valid at ray caustics, as is true of any pure ray approach. However, from experience with both the SBR and GRE, the effects of errors resulting from the proximity of ray caustics to the points of reflection and to the plane of integration at the aperture are not serious; such ray tubes are small in number and hence are ignorable as they carry only a small fraction of the total power. In the GRE method, the incident fields in the aperture at the open end are first replaced with equivalent surface currents (Huygen's sources) which radiate the equivalent coupled fields into the cavity interior. Next, the open end, or aperture, is broken up into a relatively small number of subapertures, and the fields coupled into the cavity are found by launching a dense grid of ray tubes in all directions into the cavity from the phase center of each subaperture, and then tracking them within the cavity through multiple reflections from the interior cavity walls using the laws of GO, as illustrated in fig. 3. Each ray tube is amplitude weighted according to the far zone radiation pattern of its respective subaperture but with the cavity walls absent. Like the GB shooting method, the GRE method intrinsically includes the fields diffracted into the cavity by the edges at the open end; this results from the boundary points of the surface integrations (over the equivalent Huygen's sources in the open end) used to find the far zone patterns of all the subapertures. Also like the GB

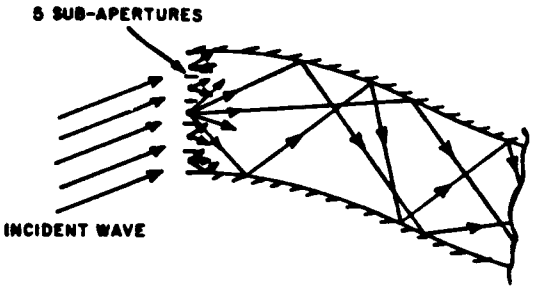


Fig. 3. Ray tubes launched from subapertures as in the GRE method.

method, the grid of ray tubes in the GRE method needs to be tracked only once independent of the incidence angles, although a much larger number of ray tubes must be tracked than the corresponding number of beams in the GB method.

While both the GRE and SBR methods are ray tracking intensive, the basic difference between the GRE method and the SBR method is in the way the ray-tubes are launched. The SBR method launches the ray tubes in such a way that only the GO incident field which is intercepted by the open end is tracked within the cavity, whereas the ray tubes of the GRE method are launched within the cavity in such a way that the fields diffracted into the cavity by the edges at the open end are intrinsically included. It is noted that both methods use the *laws* of GO to track the ray tubes within the cavity once they have been launched, and the aperture integration method is used to find $\boldsymbol{E}^s_{cav}$ from the ray fields, as discussed above in the paragraph on SBR. Also it is important to note that the set of ray tubes in the GRE method is tracked only once for a wide range of incidence angles while a new set of ray tubes in the SBR method must be tracked or for each new incidence angle. However, a much larger number of ray tubes must be tracked in the GRE method, in general, than the number of ray tubes for a *single* incidence angle in the SBR method. It is found that the SBR method is useful for analyzing the scattering from cavities which are very large electrically, in which the fields diffracted into the cavity by the edges at the open end are negligible in comparison with the GO fields, for a relatively small number of discrete incidence angles. The GRE method compliments the SBR method in that it is useful for analyzing the scattering from moderately large cavities in which the interior diffracted fields are not negligible, and for cases where a large number of incidence angles are needed. It is noted that the GRE method can be used for very large cavities, and is generally more efficient than SBR for a large number of incidence angles; however, in practice computer memory requirements may become excessively large because the ray data from all the rays in the GRE must be stored for later use as the incidence angle is changed. This problem can be overcome by using auxiliary data storage units (such as tape and hard disk drives or auxiliary RAM) to store the ray data.

The SBR and GRE methods will be described in more detail in sections 2 and 3 of this paper, respectively, and numerical comparisons will be presented in section 4. The appendix briefly describes the method of finding $\boldsymbol{E}^s_{rim}$.

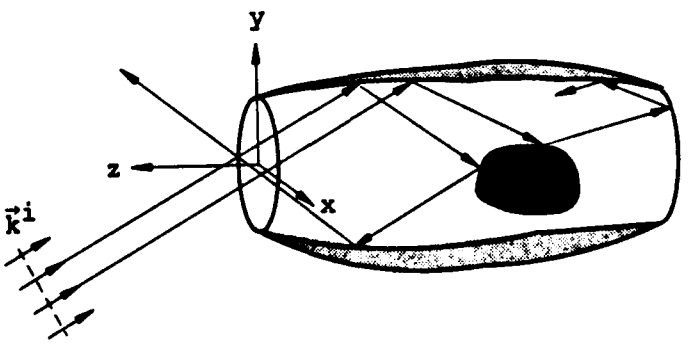


Fig. 4. SBR method for computing internal irradiation backscattering from open-ended cavities.

2. The shooting and bouncing ray method

In the SBR method, an incident plane wave is modeled by a set of parallel geometric optic rays. The rays are started or "launched" from a plane orthogonal to the propagation vector k^i of the incident wave as shown in fig. 4. Since this plane is an equi-phase surface of the incident plane wave, it is chosen to be the zero phase reference for all the rays. Generally, a regularly spaced grid of rays are launched from this plane. The actual density of the rays required is dependent on the complexity of the geometry of the cavity. For more arbitrarily shaped cavities, 10–20 rays per linear wavelength dimension is required for a convergent backscattering solution. A ray launched from the starting plane that enters the cavity is tracked as it bounces within the cavity until the ray exits the cavity through the aperture. The complex field of the ray is computed as it bounces using the laws of geometric optics. This involves calculating the ray tube divergence factors and reflection coefficients. After tracing all the rays that enter the cavity, the backscattered field and RCS is then computed from the sum of the contributions from each individual ray.

As stated earlier the advantage of the SBR method over previous methods based on modal analysis is that there is virtually no restrictions placed on the shape of the cavity. Because this method is based on high frequency methods, though, the cavity and its detailed features must be large compared to the wavelength. The cavity walls are assumed to be perfectly conducting. Thin layers of material coating may be included on the internal cavity walls. Complex terminations as well as other conducting interior scatterers may be included. Most implementations of SBR use piece-wise analytic functions to describe the complete cavity including the termination. However, numerical descriptions may also be used as long as there are accurate interpolation methods available to find the intersection between the rays and the cavity as well as to find derivatives up to second order at these intersection points.

Because the mathematical details of the ray tracing can be found in ref. [12], only a general discussion of the ray tracing procedure will be discussed here. The ray tracing of each ray involves, first, finding the intersection of each ray with the cavity. For complex cavities, numerical root finding methods must be used. This step is the most computationally time consuming step of the SBR, generally comprising 60% to 90% of the total computation time. The ray is then reflected inside the cavity at this intersection point. The direction vector of the reflected ray is determined, and this procedure is then repeated until the ray exits from the cavity.

As the ray is tracked within the cavity, the field amplitude of the ray is computed at each intersection point using geometric optics. Because it is necessary to compute the fields only at the intersection points and not along every point along the rays, the well known problem of the GO field being singular at the GO ray caustics is generally avoided. The latter problem would occur only when a ray caustic occurs right at or extremely close to an intersection or interior surface reflection point. From experience, this has been found to be extremely rare. Should such a situation occur, this ray can be discarded without noticeably affecting the final RCS result if a sufficiently large number of rays are launched. In geometrical optics, associated with each ray is a differential ray tube that can converge or diverge as the ray propagates. The electric field $E(x_i, y_i, z_i)$ at a reflection point (x_i, y_i, z_i) immediately after the reflection is found iteratively from the field at the previous reflection point from

$$E(x_i, y_i, z_i) = (\Gamma) \cdot E(x_{i-1}, y_{i-1}, z_{i-1})(\mathrm{DF})\, e^{-jks}, \quad (2)$$

where k is the wavenumber in the propagation medium which is assumed here to be free space in which the cavity is embedded, s is the distance between the consecutive points of reflection, and $E(x_{i-1}, y_{i-1}, z_{i-1})$ is the electric field at the previous point $(x_{i-1}, y_{i-1}, z_{i-1})$. (Γ) is the planar reflection coefficient matrix at the reflection point where the original curved surface is replaced by its local tangent plane at that point and (DF) is the divergence factor which governs the spreading of the differential ray tube after reflection.

The well known Fresnel (planar) reflection coefficients are used and can be found in most elementary texts on electromagnetics. Since the reflection coefficients of the electric field are different for its components parallel and perpendicular to the plane of incidence the electric field incident at the intersection (reflection) point must therefore be decomposed into its parallel and perpendicular components.

The divergence factor ((DF) in eq. (2)) is computed from the geometric optics formula

$$(\mathrm{DF}) = \frac{1}{\sqrt{1+s/R_1}}\,\frac{1}{\sqrt{1+s/R_2}}, \tag{3}$$

where R_1 and R_2 are the two principal radii of curvature of the reflected ray tube at the previous intersection point and s is as defined for eq. (2). Using the Q-matrix formulation of Deschamps [13], the two principal radii of curvature (R_1, R_2) can be found from the elements of a 2×2 curvature matrix called the $\mathbf{Q}$ matrix,

$$\frac{1}{R_{1,2}} = \tfrac{1}{2}\Big\{(Q_{11}+Q_{22}) \pm \sqrt{(Q_{11}+Q_{22})^2 - 4(Q_{11}Q_{22}-Q_{12}Q_{21})}\Big\}, \tag{4}$$

where Q_{mn} is the (m, n)th element of $\mathbf{Q}$. The ray associated with a plane wave is neither coverging nor diverging and has a curvature matrix given by

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{5}$$

A ray with a perfect circular wavefront with radii of curvature R_0 has a $\mathbf{Q}$ matrix of

$$\mathbf{Q} = \begin{bmatrix} \dfrac{1}{R_0} & 0 \\ 0 & \dfrac{1}{R_0} \end{bmatrix}. \tag{6}$$

In the SBR method, all rays initially have the zero matrix as their curvature matrix at their starting points on the initial orthogonal plane. The curvatur matrix propagates along a ray given by the relation

$$\mathbf{Q}^b = \left\{[\mathbf{Q}^a]^{-1} + s\mathbf{I}\right\}^{-1}, \tag{7}$$

where a and b are the two points along the ray (a precedes b along the ray), $\mathbf{Q}^a$ and $\mathbf{Q}^b$ are the curvature matrices at points a and b, respectively, s is the distance between a and b, and $\mathbf{I}$ is the unit dyad. Note that for the rays associated with a plane wave, the curvature matrix remains the zero matrix as the rays propagate.

After a ray is reflected from a non-planar surface, the wavefront of the ray will no longer be planar and the curvature matrix will no longer be zero. The curvature matrix for a ray right after a reflection $\mathbf{Q}^r$ is related to the curvature matrix of the ray right before the reflection $\mathbf{Q}^i$ and the curvature matix of the surface at the intersection point $\mathbf{Q}^\Sigma$ by enforcing the phase matching condition of the incident and reflected wavefronts at the reflection surface. As mentioned earlier, the mathematical details of the above described procedure can be found in ref. [12].

2.1. Aperture integration

All of the rays that enter the cavity are tracked within the cavity, as described in the previous section, until they return back to the open end. The scattered field can then be found from the standard physical-optics approximation by performing an integration over the equivalent sources in the aperture. Using a magnetic current formulation, the aperture field is replaced with an equivalent magnetic current sheet $\boldsymbol{K}_S$ whose value is approximated as follows:

$$\boldsymbol{K}_S = \begin{cases} 2\boldsymbol{E}(x_N, y_N, 0) \times \hat{\boldsymbol{n}}, & \text{over the aperture,} \\ 0, & \text{outside the aperture.} \end{cases} \tag{8}$$

The above approximation is essentially of the Kirchhoff type which employs only equivalent magnetic currents instead of both electric and magnetic currents. Hence, the strength of the magnetic current is doubled as evidenced by the factor of 2 present in eq. (8). An improvement to

the Kirchhoff type approximation in eq. (8) would be to include Ufimtsev equivalent edge currents (as in the PTD) [8]. However, the latter effects are small for electrically large apertures and for aspect angles not close to grazing the aperture plane, and are thus ignored. For a cavity with its aperture on the $z = 0$ plane and the open end pointing in the $+\hat{z}$ direction as shown in fig. 4, $\hat{n} = \hat{z}$ in eq. (8). The cavity backscattered field can be found from:

$$E^{bs} = \frac{e^{-jk_0 r}}{r}\left[\hat{\theta}^i A_\theta + \hat{\phi}^i A_\phi\right] \quad \text{as} \quad k_0 r \to \infty,$$

$$\begin{bmatrix} A_\theta \\ A_\phi \end{bmatrix} = \frac{jk_0}{2\pi}\iint_{\Sigma_A} dx\, dy\, e^{jk_0(ux+vy)} \times \begin{bmatrix} E_x \cos\phi^i + E_y \sin\phi^i \\ \left(-E_x \sin\phi^i + E_y \cos\phi^i\right)\cos\theta^i \end{bmatrix}, \tag{9}$$

$$u = \sin\theta^i \cos\phi^i, \qquad v = \sin\theta^i \sin\phi^i.$$

E_x and E_y are the x and y components of the field exiting the cavity aperture.

In reflector antenna problems, the radiated far fields are computed by performing a similar integral over the reflector aperture plane. For those problems, the points in which the aperture fields are known are generally over a regularly spaced grid, and such methods as the fast fourier transform (FFT) can be used to perform the integration in eq. (9) [14]. Even if the points were not all uniformly distributed, the fields in the aperture of a reflector antenna are smoothly varying and can be easily interpolated. For the cavity problem, however, the position of the rays over the aperture plane will not be uniformly distributed and interpolation schemes are difficult. Instead, the integration is performed discretely over each ray tube, and the total scattered fields is found from the superposition of the scattered fields of all the ray tubes. Additionally, if the output ray tube is small compared to the wavelength, the phase change across the ray tube can be approximated by a linear phase variation. For a ray striking (or exiting) the aperture at (x_i, y_i) with direction vector (s_x, s_y, s_z), the field of the ray tube across the aperture can be approximated by

$$\begin{bmatrix} E_x(x, y) \\ E_y(x, y) \end{bmatrix} = \begin{bmatrix} E_x(x_i, y_i) \\ E_y(x_i, y_i) \end{bmatrix} e^{jk_0[s_x(x-x_i)+s_y(y-y_i)]}. \tag{10}$$

Substituting eq. (10) into eq. (9) and noting that for each ray tube, $E_x(x_i, y_i)$ and $E_y(x_i, y_i)$ are independent of the integration variables x and y, the physical optics integral of eq. (9) can be written as a summation over all the ray tubes:

$$\begin{bmatrix} A_\theta \\ A_\phi \end{bmatrix} = \frac{jk_0}{2\pi} \times \sum_{\substack{i \\ \text{all rays}}} \begin{bmatrix} E_x(x_i, y_i)\cos\phi^i + E_y(x_i, y_i)\sin\phi^i \\ -E_x(x_i, y_i)\sin\phi^i\cos\theta^i + E_y(x_i, y_i)\cos\phi^i\cos\theta^i \end{bmatrix} \times e^{jk_0[s_x x_i + s_y y_i]}(\Delta x_i\, \Delta y_i) I_i, \tag{11}$$

where

$$I_i = \frac{1}{(\Delta x_i\, \Delta y_i)}\int_{\substack{i\text{th} \\ \text{ray}}}\int_{\substack{\text{exit} \\ \text{tube}}} dx\, dy\, e^{jk_0[(u-s_x)x+(v-s_y)y]}$$

and

$(\Delta x_i\, \Delta y_i)$ = area of the exit ray tube.

I_i is the normalized Fourier transform of the ray tube shape function normalized with respect to the ray tube area. Methods to compute this transform accurately by shooting auxiliary rays are described in ref. [12]. It has been found however, that if all the exiting ray tubes are small compared to the wavelength, then the approximation

$$I_i = \frac{\sin(u - s_x)}{u - s_x}\,\frac{\sin(v - s_y)}{v - s_y} \tag{12}$$

is found to be quite accurate.

Finally, the output ray tube area $(\Delta x_i\, \Delta y_i)$ can be found by using the ray tube divergence factors calculated earlier. The area of the ray tube as a ray exits is related to the area of the incident ray

tube $(\Delta x_0, \Delta y_0)$ via the product of the divergence factors:

$$(\Delta x_i, \Delta y_i) = \left[\prod_{i=1}^{N} |(\mathrm{DF})_i|\right]^{-2} \left(\frac{\cos\theta^i}{\cos\theta^e}\right)(\Delta x_0\, \Delta y_0). \quad (13)$$

The ratio of the cosines accounts for the oblique angles of the ray tubes entering and exiting the cavity. θ^i is the angle formed by the incident direction vector of the ray and the normal to the integration plane while θ^e is the angle formed by the exiting direction vector and the normal to the integration plane.

3. The generalized ray expansion method

As mentioned in the introduction, the basic difference between the SBR method and the GRE method as presented here, is in the way the ray tubes are initially launched inside the cavity. The procedures for tracking the ray tubes inside the cavity using the laws of GO and for obtaining the cavity scattered field $\boldsymbol{E}^s_{cav}$ via aperture integration are identical in both approaches. These procedures were described in detail in the last section and will not be repeated here.

In the GRE method, the incident field in the aperture of the cavity is replaced by equivalent surface currents which radiate the desired fields from the aperture into the cavity, as shown in fig. 5. The exact aperture fields can be used if they can be found; however, for sufficiently large apertures, the fields launched into the cavity by the incident wave can be found via a Kirchhoff approximation for the aperture field. This is reasonable for incidence angles which are not close to grazing along the aperture surface. The equivalent electric and magnetic surface currents in the aperture Σ_A are then given by

$$\boldsymbol{J}^{eq}_s(\boldsymbol{r}') = -\hat{\boldsymbol{n}} \times \boldsymbol{H}^i(\boldsymbol{r}'), \qquad \boldsymbol{M}^{eq}_s(\boldsymbol{r}') = -\boldsymbol{E}^i(\boldsymbol{r}') \times \hat{\boldsymbol{n}}, \quad (14)$$

where $\boldsymbol{E}^i$ and $\boldsymbol{H}^i$ are the incident electric and magnetic fields, respectively, $\boldsymbol{r}'$ is the vector from the origin O to any point in Σ_A, and $\hat{\boldsymbol{n}}$ is again

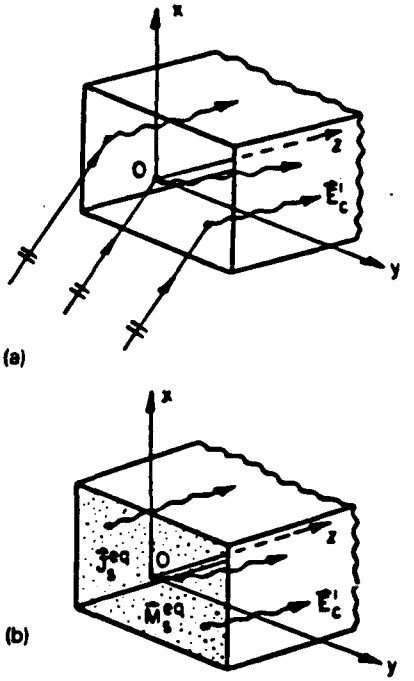


Fig. 5. Coupling of the incident plane wave into the cavity using equivalent surface currents in the aperture (open end). (a) Original problem; (b) equivalent problem.

the unit surface normal of Σ_A at the point $\boldsymbol{r}'$, pointing out of the cavity.

According to Huygen's principle, each infinitesimal current element (or Huygen's source) defined by $\boldsymbol{J}^{eq}_s(\boldsymbol{r}')\,ds'$ and $\boldsymbol{M}^{eq}_s(\boldsymbol{r}')\,ds'$, where ds' is an infinitesimal surface area at $\boldsymbol{r}' \in \Sigma_A$, launches spherical waves which can be tracked within the cavity via the laws of GO. However, in practical applications it is useful to deal with a discrete array of a small number of *finite* size sources which also launch waves that become spherical at a reasonably short distance away, rather than to deal with a continuum of infinitesimal sources. Hence, the original aperture Σ_A is divided into a relatively small number of subapertures as in fig. 6. Since the aperture in fig. 6 is rectangular, it is convenient to use rectangular subapertures (non-rectangular apertures can be subdivided, for example, into polygonal subapertures); in particular,

$$\Sigma_A = \sum_{l=1}^{\mathscr{L}} S_l, \quad (15)$$

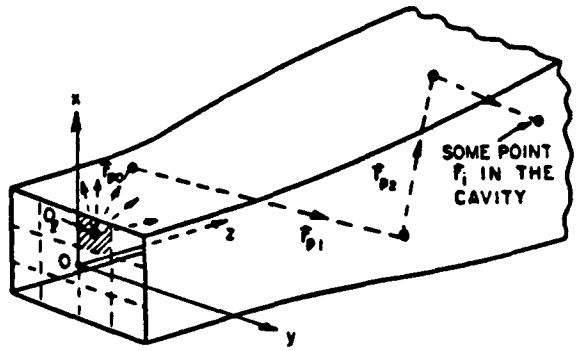


Fig. 6. Ray field originating from the phase center O_l of the lth subaperture of area S_l (shaded).

where the size of the typical lth subaperture area S_l is determined to a large extent by the overall length of the cavity and the size of the original aperture Σ_A. In general, the maximum linear dimension D of the largest subaperture should be less then $\sqrt{2L\lambda}$ where L is the overall length of the waveguide cavity and λ is the free space wavelength. Furthermore, the number of subapertures $\mathscr{L}$ should be greater than one. Otherwise, the subaperture grid can be chosen arbitrarily for convenience; however, for efficiency the smallest possible number of subapertures should be used which yields an adequately convergent solution.

With an appropriate choice of subaperture discretization, one can then show that the field radiated into the cavity by the equivalent current elements distributed over each of the subapertures can be well approximated by a superposition of an array of spherical waves originating from the phase center of each of the $\mathscr{L}$ subapertures. These waves can then be tracked along their associated rays which undergo successive reflections within the cavity as in fig. 6, just as in the previously mentioned conventional ray shooting GO (SBR) approach. Furthermore, the *initial* amplitudes of the rays emanating in all directions (in the half-space forward of the subaperture) from the phase center of each subaperture are directly given by the radiation pattern of that subaperture with the cavity walls absent. The ray field originating from the phase center O_l of each subaperture is broken up into a dense grid of ray tubes which initially have spherical wavefronts (before they reflect from the curved walls of the cavity), as in fig. 7. It is noted from fig. 7 that it is usually not necessary to launch ray tubes in all directions in the half-space forward of the subaperture. In practice, the ray tubes are limited to a cone with a half-angle less than 90° and approximately 15° greater than the maximum angle the incident plane wave makes with the cavity axis. The electric field along the pth ray tube launched from the phase center O_l of the lth subaperture, before the ray tube encounters a wall, is then given by

$$E_{pl}(r_{pl}) \sim C_l(\hat{r}_{pl})\frac{e^{-jkr_{pl}}}{r_{pl}}, \tag{16}$$

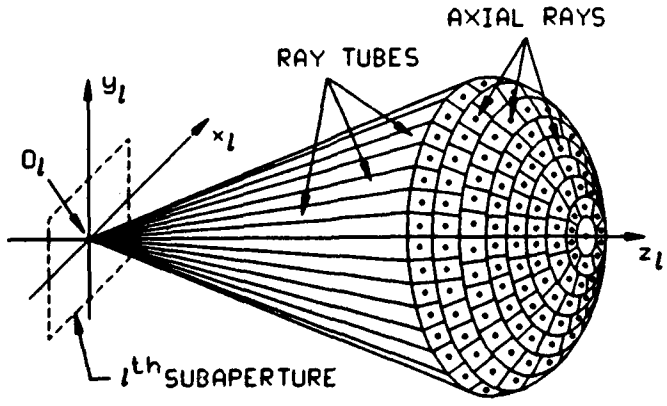


Fig. 7. Grid of ray tubes originating from O_l of any subaperture of area S_l.

where $r_{pl} = \hat{r}_{pl} r_{pl}$ is the vector from O_l to a point along the pth ray tube before its first reflection. $C_l(\hat{r}_{pl})$ is the far zone vector radiation pattern of the electric field evaluated in the direction $\hat{r}_{pl}$ with the cavity walls absent, and is thus given by

$$C_l(\hat{r}_{pl}) = \frac{jkZ_0}{4\pi} \iint_{S_l} \left[\hat{r}_{pl} \times \hat{r}_{pl} \times J_s^{eq}(r_l') + Y_0 \hat{r}_{pl} \times M_s^{eq}(\hat{r}_l') \right] e^{jk\hat{r}_{pl} \cdot r_l'} \, ds', \quad (17)$$

where r_l' is a vector from O_l to the sources at any point in S_l. Equation (17) can be evaluated numerically, but if the lth subaperture is a polygon and the incident field is a plane wave, then eq. (17) can be evaluated in closed form. This is because the Fourier transform of a constant amplitude, linearly phased function existing over a polygonal area is known [15], and eq. (17) reduces to such a transform for the case of plane wave incidence.

As shown in fig. 6, the pth ray originating from O_l is tracked to some interior point at r_i. This ray tracking is performed in exactly the same manner as in the SBR method described in section 2. All of the ray tubes are tracked within the cavity until they return back to the open end. The field E_{cav}^s is then found by performing an aperture integration over the equivalent sources defined by the projection of each ray tube in the plane of the aperture Σ_A as the ray tube exits the cavity. Again, this is exactly the same procedure as in the SBR method.

As indicated by eqs. (16) and (17), only the initial amplitude coefficient $C_l(\hat{r}_{pl})$ of the plth ray tube depends on the incident field, and furthermore, this coefficient is found given only the initial direction $\hat{r}_{pl}$ of the ray tube. Therefore, the ice cream cone-shaped grid (fig. 7) of ray tubes of each subaperture can be launched with unit amplitudes, tracked throughout the cavity and the corresponding ray data stored. Then, as the incident field changes, this ray data can be amplitude weighted according to eqs. (16) and (17), without having to re-track the rays.

It is noted that a large number of ray tubes need to be tracked in the GRE, albeit only once, for a given waveguide cavity. In contrast, the SBR approach requires fewer rays to be tracked for each incidence angle, even though the total number of rays tracked in the SBR is much larger than those in the GRE when considering a wide range of incidence angles. Therefore, the SBR appears to be more efficient if only a few (say ten or less) incidence angles are required in the calculations, whereas the GRE is more efficient if several incidence angles or a continuous range of incidence angles are necessary.

4. Numerical results

The numerical results presented here are in the form of rader cross-section (RCS) vs. θ, where θ is the angle between the propagation vector of the incident plane wave k^i and the waveguide cavity axis. RCS is defined by

$$\sigma = \lim_{r \to \infty} 4\pi r^2 \frac{|E^s|^2}{|E^i|^2}, \quad (18)$$

where r is the distance from the open end of the cavity to a far field observer, and $|E^i|$ and $|E^s|$ are the magnitudes of the incident and scattered fields, respectively. The units of RCS are given in "decibels relative to a square wavelength" (DBSW) (i.e. $10 \log_{10} \sigma$ with σ in square wavelengths). The ray tracing subroutines used in the codes to generate the numerical results for both the SBR and GRE methods are adapted from codes developed by Prof. S.W. Lee at the University of Illinois, and are based on a super-elliptic geometry model.

Figure 8 shows the RCS patterns of a relatively small rectangular cavity, with the interior cavity scattering calculated using the SBR and GRE methods and compared with a reference solution based on the hybrid modal method. Perpendicular and parallel polarization in all the figures refer to the incident electric field vector being either perpendicular or parallel to the plane defined by the direction vector of the incident plane wave and the axis of the cylinder (the z-axis). Both of the ray-based methods agree reasonably well with the reference solution, which is somewhat surprising considering the high fre-

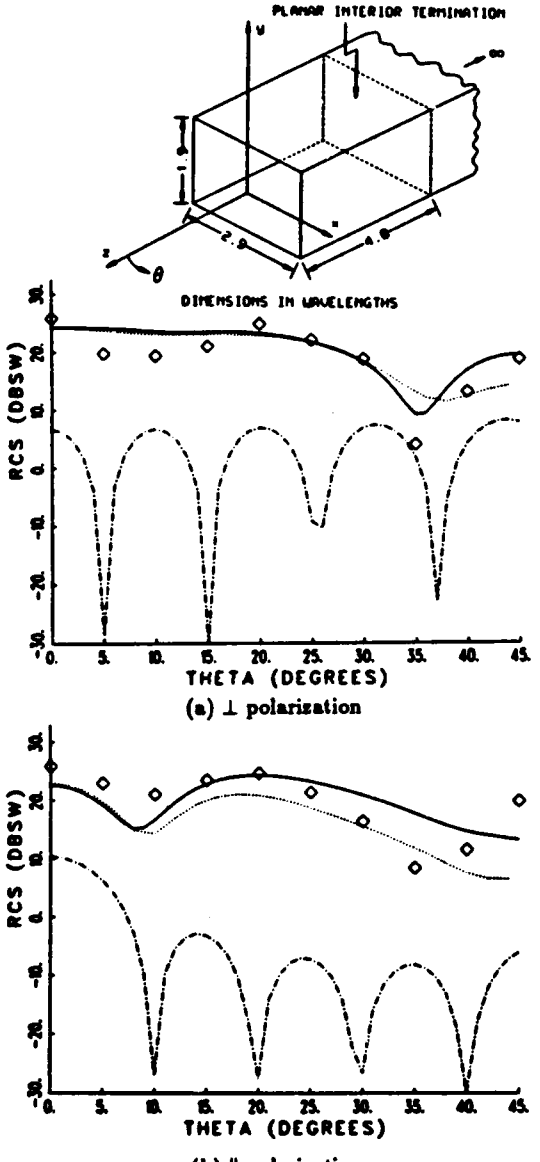


Fig. 8. RCS patterns in the $x-z$ plane of a small rectangular cavity calculated using the GRE, SBR and hybrid modal (reference solution) methods. The scattering from the rim is calculated using the equivalent current method. ——— hybrid modal; ····· GRE; ◇ SBR; –·– rim.

quency asymptotic approximations of ray tracing methods. The scattering from just the rim at the open end is also shown in this figure, and was found using the equivalent current method (ECM) described in the appendix. Because the SBR method requires that a new set of rays be tracked for each incidence angle, the SBR results are calculated at 5 degree increments in this figure (represented by the diamond shaped markers). The GRE results are calculated at much finer increments because only one set of rays is tracked for the entire range of incidence angles, and the scattered field for each incidence angle is calculated very quickly from the stored ray data. It is noted that in both of these ray-based methods, the ray tracking typically requires more than 80% of the total CPU time used in the calculations, and for the data plotted in fig. 8, both methods used approximately the same total number of rays.

In the remainder of the numerical results presented here, only the scattering component due to the cavity interior will be shown, i.e. no external scattering effects are included, such as the scattering from the rim. Figures 9 and 10 show the RCS patterns of a larger rectangular cavity and an annular cavity with a retangular cross-section, again comparing GRE and SBR with a modal reference solution. The GRE and SBR methods agree quite well with the reference solution in the case of the rectangular cavity, and reasonably well in the case of the annular cavity. It is expected that the two ray-based methods will have slightly more error when used to analyze cavities with curved interior walls because they cannot account for curved surface diffraction. However, this is a relatively weak effect, especially for electrically large cavities.

Figure 11 shows the RCS patterns of a large cavity which transitions from a rectangular cross-section at the open end to a circular termination and the axis of the cavity follows third-order polynomial "S-bend". No modal reference solution is available for this geometry, so only the GRE and SBR results are shown, and they agree with each other reasonably well.

Comparisons of the SBR solution and the modal solution for the RCS from a large circular cylindrical cavity can be found in ref. [12]. A GRE solution is not available for this cavity at the present time because of the extra algorithm development necessary to subdivide a circular open end into appropriate subaperture domains;

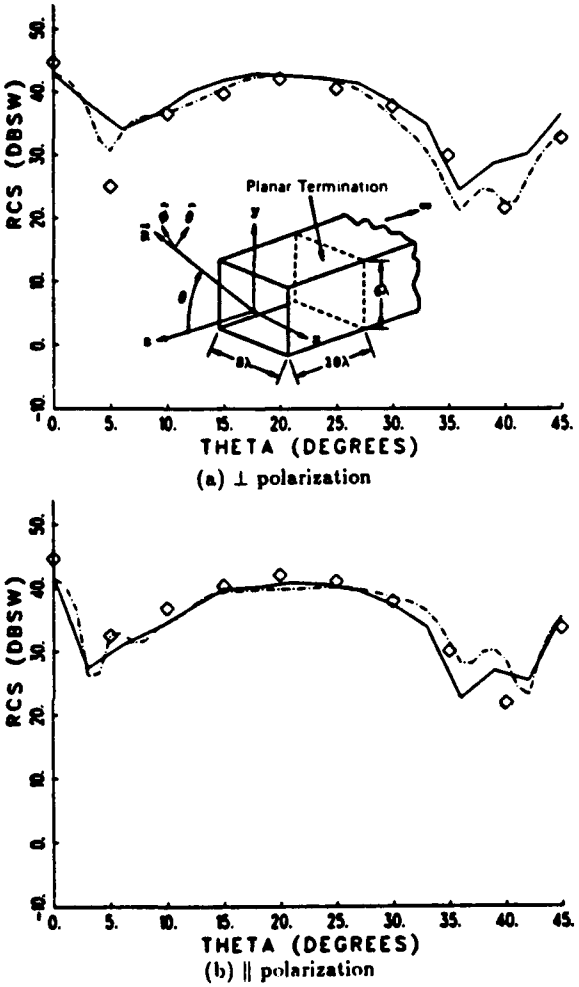


Fig. 9. RCS patterns in the $x-z$ plane of a larger rectangular cavity calculated using the GRE, SBR and hybrid modal (reference solution) methods. ——— hybrid modal; -·- GRE; ◇ SBR.

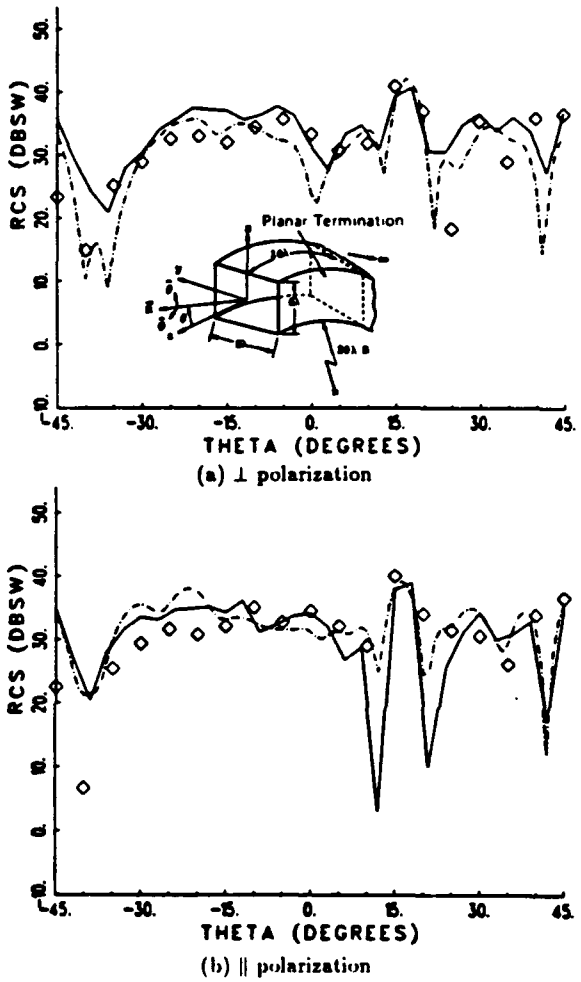


Fig. 10. RCS patterns in the $y-z$ plane of a large annular cavity with a rectangular cross-section calculated using the GRE, SBR and hybrid modal (reference solution) methods. ——— hybrid modal; -·- GRE; ◇ SBR.

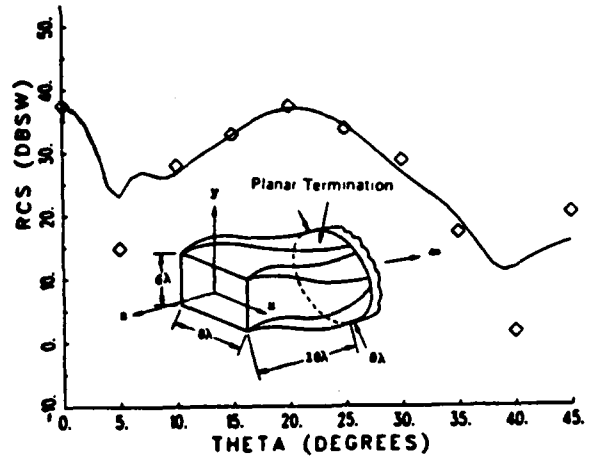
however, such an algorithm should be available in the near future. All of the GRE results presented here have been for cavities with rectangular open ends, which are very easy to subdivide into a number of rectangular subapertures. However, the GRE method can in principle be applied to cavities with fairly arbitrarily shaped, non-planar open ends in a straightforward manner. As mentioned above, work is currently under way to develop subaperture gridding algorithms for arbitrarily shaped open ends which require a minimum of user interaction.

Appendix

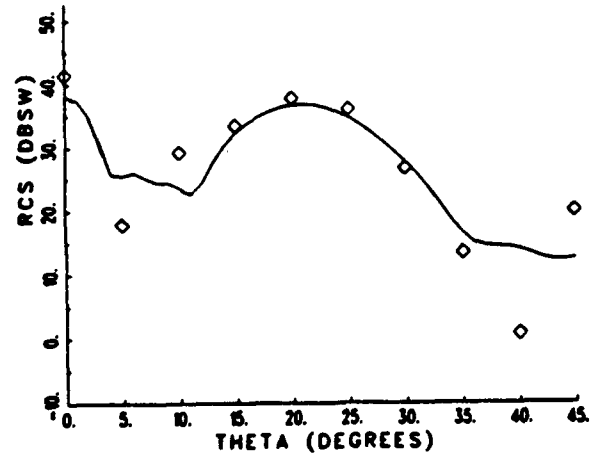
The contribution to the external scattering from just the rim of an open-ended waveguide cavity is found via the ECM as follows:

$$E_0^s \approx \frac{jkZ_0}{4\pi}\frac{e^{-jkr}}{r}\int_{\text{rim}}\left[\hat{r}\times\hat{r}\times I_{\text{eq}}(t')\hat{t}' + Y_0\hat{r}\times M_{\text{eq}}(t')\hat{t}'\right]e^{jk\hat{r}\cdot r'}\,dt', \qquad (19)$$

in which r denotes the position vector to the far zone observation point from a conveniently cho-



(a) ⊥ polarization



(b) ∥ polarization

Fig. 11. RCS patterns in the $x-z$ plane of a large "offset" cavity with a cross-section which transitions from rectangular at the open end to a circular termination calculated using the GRE and SBR methods. The offset is 3λ in the $-\hat{y}$ direction. —— GRE; ◇ SBR.

sen coordinate origin (at O), Z_0 and Y_0 are the free space impedance and admittance, respectively, and the integration over the equivalent electric (I_{eq}) and magnetic (M_{eq}) current sources is performed along the rim or the contour formed by the aperture edge at the open end. These equivalent currents can be found via GTD in terms of the fields incident on the edge and the appropriate diffraction coefficients as is done in ref. [2].

References

[1] H. Ling, R.-C. Chou and S.W. Lee, IEEE Trans. Antennas Propag. 37 (1989) 194.
[2] P.H. Pathak and R.J. Burkholder, IEEE Trans. Antennas Propag. 37 (1989) 635.
[3] R.J. Burkholder, High-Frequency Asymptotic Methods for Analyzing the EM Scattering by Open-Ended Waveguide Cavities, Ph.D. dissertation, The Ohio State University, Columbus (1989).
[4] P.H. Pathak and R.J. Burkholder, Radio Sci. 26 (1991) 211.
[5] J.B. Keller, J. Opt. Soc. Am. 52 (1962) 116.
[6] R.G. Kouyoumjian and P.H. Pathak, in: Acousic, EM and Elastic Wave Scattering, V.K. Varadan and V.V. Varadan, eds. (North-Holland, Amsterdam, 1986) p. 265.
[7] P.H. Pathak, Antenna Handbook – Theory, Applications, and Design (Van Nostrand Reinhold, New York, 1988) Ch. 4.
[8] P.Ya. Ufimtsev, Izd-Vo Sov. Radio (1962) 1 (in Russian) (Translation prepared by the U.S. Air Force Foreign Technology Division, Wright-Patterson Air Force Base, Ohio, released for public distribution 7 September 1971).
[9] H. Shirai and L.B. Felsen, Wave Motion 9 (1987) 301.
[10] H. Shirai and L.B. Felsen, Wave Motion 9 (1987) 461.
[11] J. Maciel and L.B. Felsen, IEEE Trans. Antennas Propag. 38(10) (1990) 1607.
[12] R. Chou, Reduction of the Radar Cross Section of Arbitrarily Shaped Cavity Structures, Ph.D. dissertation, University of Illinois, Champaign (1987).
[13] G.A. Deschamps, Proc. IEEE 60 (1972) 1022.
[14] Y. Rahmat-Samii, Antenna Handbook – Theory, Application, and Design (Van Nostrand Reinhold, New York, 1988) ch. 15.
[15] S.W. Lee and R. Mittra, IEEE Trans. Antennas Propag. 31 (1983) 99.

Analysis of EM Penetration into and Scattering by Electrically Large Open Waveguide Cavities Using Gaussian Beam Shooting

ROBERT J. BURKHOLDER, MEMBER, IEEE, AND PRABHAKAR H. PATHAK, FELLOW, IEEE

An approximate Gaussian beam (GB) shooting method is presented for analyzing the electromagnetic coupling into and scattering by electrically large perfectly conducting, open-ended, nonuniform waveguide cavities with a slowly varying wall curvature. The method is illustrated only for the two-dimensional (2-D) case; however, the ideas developed here can be extended to deal with the corresponding three-dimensional (3-D) situation. An external plane wave directly illuminates the aperture (open end) and the incident field in the aperture at the open end is found via the Kirchhoff approximation. This aperture field is then expanded into a discrete set of identical GB's which are launched radially into the cavity from an array of preselected points in the aperture, and this array of points corresponds to the phase centers of the subapertures into which the aperture is divided. The initial beam launching amplitudes are found by matching GB's to the radiation field of each subaperture. The GB's thus launched from the aperture propagate into the waveguide cavity and are tracked to the interior termination via multiple reflections at the cavity walls. For the sake of efficiency and simplicity, the GB's are tracked axially in this paper using the rules of beam optics which ignore any beam distortion upon reflection at the walls. The effects of beam distortion are not significant for relatively slowly varying waveguide cavities. Finally, the field scattered into the exterior by the termination within the cavity is found using a reciprocity integral formulation which requires a knowledge of the beam fields near the termination. Numerical results based on this GB approach are presented and compared with results based on an independent reference solution.

I. INTRODUCTION

The problem of representing high frequency electromagnetic (EM) fields which propagate within closed waveguides of relatively arbitrary shape (for which modes cannot be defined in the conventional sense) is a rather complex one. Its solution is crucial for dealing with the important problem of EM scattering from (as well as interior coupling into) an electrically large open-ended waveguide cavity of relatively arbitrary shape which contains an interior termination, when it is illuminated by an external source as depicted in Fig. 1. In this paper, the configuration of Fig. 1 is surrounded by free-space and has perfectly conducting interior cavity walls which may have a thin material coating. It is analyzed by employing a set of GB field basis functions to represent and track the fields which are coupled from the incident wave, via the open front end, into the interior waveguide cavity region. It is assumed in the present analysis that the incident field directly illuminates the aperture defined by the open front end of the cavity, and that the aspects of incidence and scattering are primarily restricted to within about 60° of being normal to the aperture, so that one may employ the Kirchhoff approximation to represent the incident fields in the electrically large aperture. Outside this angular region, the external features of the cavity will generally dominate the scattered field.

An advantage of the present approach for analyzing the configuration of Fig. 1 is that the GB's need to be tracked only once within the cavity independent of the excitation or changes in directions of incidence and scattering; it is necessary only to change the initial beam launching amplitudes with change in source illumination, because the beam launching directions are made independent of the source excitation. Furthermore, this approach implicitly takes into account, to within the Kirchhoff approximation, the contribution to the field coupled into the cavity via diffraction by the edges at the open end. The method described in this paper has been discussed briefly in [1], and is based on the work reported in [2]; these two references also describe some other useful approaches for analyzing this type of EM scattering/penetration problem.

As illustrated in Fig. 1, the high frequency field $\vec{E}^s$ scattered by an open cavity is composed basically of three

Manuscript received May 29, 1990; revised March 11, 1991. This work was supported in part by NASA Lewis Research Center under Grant NAG3-476 and by the Joint Services Electronics Program under Contract 721564 to The Ohio State University.

The authors are with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43212.

IEEE Log Number 9103311.

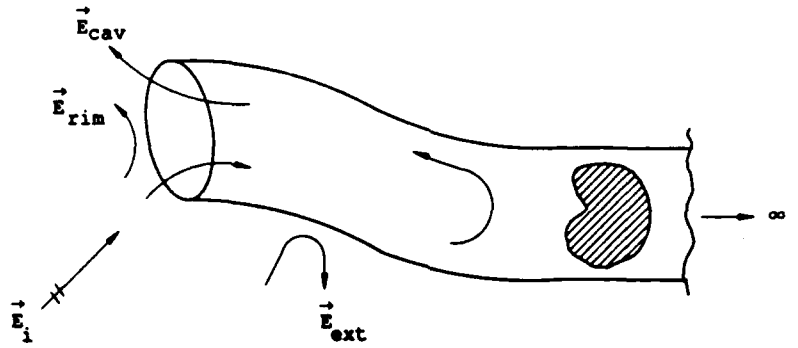


Fig. 1. Open-ended waveguide cavity and the scattering mechanisms associated with plane wave illumination.

terms:

$$\vec{E}^s = \vec{E}^{rim} + \vec{E}^{ext} + \vec{E}^{cav} \tag{1}$$

where $\vec{E}^{rim}$ is due to the scattering from the rim at the open end, $\vec{E}^{ext}$ is due to the scattering from all external features of the cavity configuration other than the rim at the open end, and $\vec{E}^{cav}$ denotes the scattering arising from the incident EM energy which couples into and propagates within the interior of the cavity before radiating out via the open end. The contribution $\vec{E}^{ext}$ is not considered here as it depends on the specific geometrical configuration which houses the cavity; however, it could in general be found via high frequency techniques such as the GTD [3] (or its uniform version, the UTD [4], [5]) or by the PTD [6], [7], [5]. $\vec{E}^{rim}$ may be found using the equivalent current method [5] combined with the GTD. Since the technique for finding $\vec{E}^{rim}$ has been described previously in some detail [1], [2], it will not be repeated here. $\vec{E}^{cav}$ is found using the GB shooting technique presented here. In general, $\vec{E}^{cav}$ can be further decomposed as:

$$\vec{E}^{cav} = \vec{E}^{term} + \vec{E}^{refl} + \vec{E}^{mult} \tag{2}$$

where $\vec{E}^{term}$ is usually the dominant contribution to $\vec{E}^{cav}$ and is due to the incident EM energy which is coupled into the cavity and then propagates to the termination from where it is reflected to radiate out from the open end. $\vec{E}^{refl}$ is due to the incident EM energy coupled into the cavity and which is reflected back without reaching the termination to then radiate out from the open end. $\vec{E}^{mult}$ describes all the higher order multiple wave interactions within the cavity not included in $\vec{E}^{term}$ and $\vec{E}^{refl}$. $\vec{E}^{mult}$ is generally negligible compared with $\vec{E}^{term}$ and/or $\vec{E}^{refl}$ for electrically large cavities and hence will be neglected here.[1] $\vec{E}^{refl}$ can be found by tracking the fields coupled into the cavity using the GB shooting method presented here, and allowing those beams which return to the open end before reaching the termination to radiate into the exterior region via the open end. $\vec{E}^{refl}$ may then be calculated via an aperture integration based again on the Kirchhoff approximation, wherein one employs the fields associated with the beams which are reflected back to the open end before reaching the termination to define equivalent sources in the aperture at the open end. It is noted that the cavity scattering contribution $\vec{E}^{refl}$ usually arises from the presence of tapers in the waveguide cavity cross section. $\vec{E}^{refl}$ will not be considered further in this paper because of space limitations, and emphasis will be placed on the dominant term $\vec{E}^{term}$.

In the present approach, the fields in the aperture at the open end which are excited by the incident wave, are expanded into a set of identical, angularly shifted GB's whose initial beam directions at launch have a constant angular interbeam spacing. Thus a set of GB's are launched radially into the cavity from each point in an array of preselected points in the aperture; i.e., all GB's in a given set have their waists centered at the point located in the array which defines that particular set. Therefore, each set of angularly tilted GB's are launched from points in the aperture which are spatially shifted with respect to the launching points of its neighboring or adjacent sets of GB's as illustrated in Fig. 2(a). This manner of launching GB's is in contrast to the method of launching a grid of parallel geometrical optics (GO) incident ray tubes into a cavity illuminated by a plane wave as in the shooting and bouncing ray (SBR) method [8], [1], [2] for analyzing the same problem. As Fig. 2(b) illustrates, the SBR method which tracks only GO rays therefore produces regions of geometrical shadows within the cavity which would otherwise have been penetrated by the fields diffracted from the edges at the open end, and this GO approach also fails near ray caustics which may be present within the cavity. On the other hand, the GB shooting method overcomes both of these difficulties present in the GO based SBR solution. Also, far fewer GB's need to be tracked within the cavity than ray tubes, in general, because a new grid of ray tubes needs to be tracked in the SBR method for each new direction of the incident plane wave. Nevertheless, it may be mentioned that the SBR method is useful in that it predicts the dominant scattering effects of open-ended cavities (Fig. 1) quite well at high frequencies.

While there is some arbitrariness in choosing the spatial shifts or locations of the array of points from which to

[1] The multiple wave interactions across the electrically large aperture are weak in comparison to the first order effects being considered here; likewise the multiple wave interactions between the open end and the termination are negligible if the interior reflection from the termination is significant since the interior reflection from the open end is generally much weaker for electrically large openings.

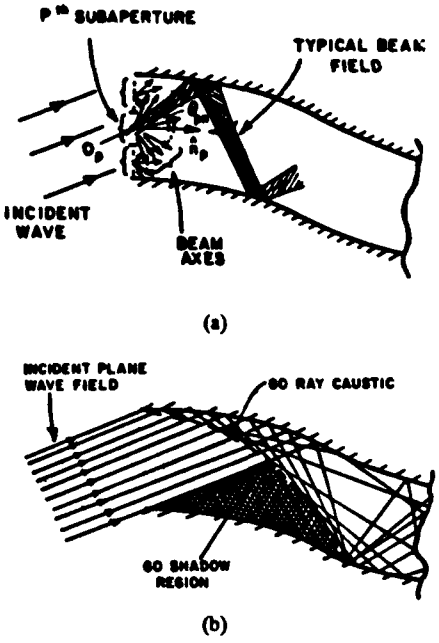


Fig. 2. GB and GO ray launching schemes for cavities. (a) The GB shooting method. (b) The geometrical optics ray shooting method.

launch the different sets of GB's, as well as in choosing the constant angular interbeam spacing between the GB's in each set, there are physical considerations which must be invoked in order to accurately and efficiently track the GB basis functions as they evolve within the cavity after being launched from the aperture at the open end. These physical considerations provide information on the choice of the initial beam parameters at launch. The initial beam parameters are obtained in a quite straight forward manner and they will be discussed briefly when references to previous related work are cited below in this section; they will also be alluded to in some detail in the next section which deals with this aspect of beam launching (or beam shooting). Furthermore, it is noted that the physical considerations which provide the conditions and guidelines for a reasonable choice of the initial GB launching parameters remove most of the arbitrariness in the selection of these parameters.

Once launched, the GB's are tracked axially from the open end to the termination region in the back of the cavity via multiple reflections at the interior cavity walls. It is assumed that interior diffraction effects are small so that only the beam reflection effects are dominant. This assumption requires that the interior walls of the cavity be slowly varying and smoothly continuous with no abrupt discontinuities in the region between the open end and the termination in which the GB's are to be tracked (90° interior corners are allowed because they introduce no diffraction effects, as in the case of a cavity with a rectangular cross section). In the present approach, each individual GB is tracked as a whole only along it's beam axis which has a real ray direction in space. This axial GB tracking approximation is employed here because of its simplicity and because it provides a physical insight into the wave propagation phenomena within the cavity. Such an approximation allows one to employ the simpler rules of beam optics which provide information on transforming an incident GB into a reflected GB upon reflection at the interior cavity wall. This real (or axial) ray tracking approximation is valid only if the GB's do not suffer appreciable distortion upon reflection. If an appreciable beam distortion occurs upon each successive reflection so that the reflected beams no longer remain Gaussian about their beam axis, then one must resort to higher order approximations to calculate beam reflections, or even require that the beams be tracked as complex rays [9]–[15]. The complex ray tracking procedure is of course more cumbersome than real ray tracking. In order for the real (axial) ray tracking approximation to be valid, the initial GB parameters must be chosen such that the GB's remain reasonably well focused as they propagate from the open end to the region near the termination. Consequently, the initial GB waists should be chosen to be reasonably wide so that the GB's remain sufficiently well focused as they propagate or evolve within the cavity; however, the initial waists cannot be chosen to be too wide as the GB's must adequately fit within the waveguide as they propagate via interior reflections, and maintain a sufficient degree of resolution. Furthermore, the spot size of the GB's must not be too large as compared to the dimensions over which the surface curvature changes or else beam distortion effects can become significant and the axial tracking approximation then tends to become less accurate. It is noted that the GB expansion coefficients and the interbeam angular spacing are based on far field patterns arising from subaperture domains or windows defined naturally around the array of points from which sets of GB's are launched as in Fig. 2(a). The initial waist size determines the width of a beam in the far field; this also then determines the array point spacing and hence the subaperture domains from which the beams are launched. Thus the size of the subaperture domains is related to the initial beam waist size. In the course of this investigation, it has been found that the initial waist size should be about one-third to one-half the width of the open end of the cavity so that the beams fit into the cavity and stay relatively well focused as they evolve within the cavity via beam reflections at the slowly varying cavity walls. The number of GB's launched (and hence the interbeam spacing at launch) depends on how well one can reproduce the subaperture far field pattern with the cavity walls absent (Kirchhoff approximation). Typically, one needs to match about 3–4 GB's per lobe of the subaperture far field pattern; this in a sense provides completeness of the initial GB expansion at launch based purely on physical considerations. This choice of initial beam waist size and the number of beams launched has been found to be adequate for waveguide cavities which are about 4 times as long as they are wide; for longer cavities it may become necessary to track the GB's via complex rays, or alternatively, one could restart and again axially track a new set of GB's after the initial set has propagated a certain distance within the cavity, so as to minimize

the effects of beam distortion upon successive reflections. However, both of the latter corrective measures, which have not been incorporated into the present analysis, will add significantly to the computational time [16], [17]. Also, a self-consistent expansion of the fields in finite apertures which is based on the Gabor representation [18], [19] has been developed recently by Felsen and his coworkers [20], [15]; this can certainly be used as an alternative procedure for beam launching. The coefficients of this expansion can found rigorously using a biorthogonality property, but in certain cases they can be found by sampling the aperture fields or the far field radiation pattern of the entire aperture, depending on the initial choice of array or lattice spacing in the expansion. An advantage of the Gabor type expansion is that it guarantees completeness *a priori* and essentially removes any arbitrariness in the choice of initial beam parameters once the lattice spacing is chosen. Since this Gabor based expansion gives rise to nonidentical GB's with nonuniform angular or interbeam spacing, the GB's resulting from this expansion cannot in general satisfy the assumptions required to track the GB's inside the cavity via the axial approximation thereby necessitating that these beams be tracked as complex rays. Clearly, this Gabor based method for beam launching, as well as complex ray tracking of these beams within the cavity, merit further attention and will form a part of the ongoing and future work on this important topic of wave propagation in nonuniform waveguides; however, it would be extremely important to develop more efficient ways, if possible, to track GB's via complex rays to make this method more practical. It is hoped that this proposed work along with the information available from the methodology developed in the present paper would provide ways to develop approximate but even more efficient, accurate and useful techniques for beam launching and tracking.

In the present analysis, the termination is assumed to be a planar, perfectly conducting surface for simplicity; whereas, it is indicated later on as to how one may handle a more complex termination. A generalized reciprocity integral defined over a conveniently chosen plane (or waveguide cavity cross-section) located sufficiently near the termination, which requires a knowledge of the fields tracked down to this termination plane from the open front end, then directly furnishes the contribution to the external scattered fields arising from the waves reflected out of the cavity by the interior termination (as $\vec{E}^{term}$) provided the termination scattering characteristics are known only over that plane. In the case of a planar termination considered here, the reciprocity integral is simply defined over the termination plane itself, and the scattering characteristics of this termination are also known on that plane. A useful property of this reciprocity integral method for calculating the fields scattered by the interior termination is that the beams need to be tracked only "one way" from the open end to the termination. Thus one can enhance the efficiency of the axial GB tracking approximation by not requiring the beams to be tracked an additional distance back again to the aperture at the open end from where they radiate into the exterior. It is noted that the GB's contributing to $\vec{E}^{refl}$ mentioned earlier are reflected back to the open end without ever reaching the plane near the termination, and therefore do not contribute to this reciprocity integral.

The analysis in this paper will be restricted to two-dimensional (2-D) geometries to illustrate the method as indicated earlier; however, the ideas developed here can be extended to deal with the (three–dimensional) 3-D case. The GB's launched in the 3-D case may be assumed to be initially rotationally symmetric and the beam polarization can be selected to match the subaperture far field patterns as in the 2-D case. The GB's will become astigmatic upon reflection from an arbitrarily curved cavity wall and the subsequent axial tracking of such beams within the cavity will have to be found via rules for reflecting astigmatic GB's within the axial approximation. Fortunately, these rules of beam reflection (i.e., beam optics) are closely analogous to the laws of GO ray optics which govern the reflection of astigmatic GO ray tubes. In the present 2-D analysis, the geometry lies in the x-z plane and TM_y refers to the case where the magnetic field is transverse to the $\hat{y}$ direction so that the electric field ($\vec{E}$) is polarized in the $\hat{y}$ direction. Likewise, TE_y refers to the case where the electric field is transverse to $\hat{y}$ so the magnetic field ($\vec{H}$) is polarized in the $\hat{y}$ direction. Also, $\vec{U} = \hat{y}U$ denotes a $\hat{y}$-directed field quantity, such that $\vec{U} = \vec{E}$ for TM_y and $\vec{U} = \vec{H}$ for TE_y. An $e^{j\omega t}$ time dependence for the fields is assumed and suppressed throughout the following analysis.

II. Gaussian Beam Launching Formulation

Consider the 2-D open cavity geometry of Fig. 3 which is excited by a plane wave. The plane wave field in the aperture at the open end of this cavity can be represented via the Kirchhoff approximation by the unperturbed incident field U^i. This aperture field U^i can be represented as a sum of spatially and angularly shifted GB's by

$$U^i(x,z)|_{z=0} = \sum_m \sum_n A_{mn}(\theta_i) B(x_{mn}, z_{mn}; b)|_{z=0} \quad (3)$$

where $A_{mn}(\theta_i)$ are the expansion coefficients and θ_i is the plane wave incidence angle. $B(x_{mn}, z_{mn}; b)$ is the 2-D GB basis function for a beam along the z_{mn} axis whose waist is located at $z_{mn} = 0$. x_{mn} and z_{mn} are the relative coordinates of the mnth shifted, rotated beam given by the coordinate transformation

$$\begin{aligned} x_{mn} &= (x - mL_x)\cos\theta_n - z\sin\theta_n \\ z_{mn} &= (x - mL_x)\sin\theta_n + z\cos\theta_n \end{aligned} \quad (4)$$

which shifts the mnth beam spatially by mL_x in the x direction and rotates it angularly by θ_n from the z axis.

The 2-D GB basis function used in this formulation (for a typical beam propagating along the z axis) is a slightly modified form of that used in [22] and is given by

$$B(x,z;b) = \sqrt{\frac{jb}{z+jb}} e^{-jkz} e^{-j\frac{1}{2}k\frac{x^2}{R(z)}} e^{-\frac{x^2}{w^2(z)}}. \quad (5)$$

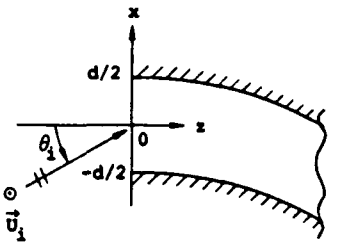


Fig. 3. 2-D open-ended cavity geometry illuminated by a plane wave.

Upon replacing x and z in (5) by x_{mn} and z_{mn} of (4), one obtains $B(x_{mn}, z_{mn}; b)$ in the expansion of (3). $R(z)$ in (5) is the usual phase front radius of curvature given by

$$R(z) = \frac{1}{z}(z^2 + b^2) \tag{6}$$

and $w(z)$ is one-half the $1/e$ beam width or "spot size" given by

$$w(z) = \sqrt{\frac{2}{kb}(z^2 + b^2)} \tag{7}$$

where b is a positive real constant which will be referred to here as the beam parameter. k is the free space wavenumber ($k = 2\pi/\lambda$), and the beam basis function of (5) is normalized such that $B(0, 0; b) = 1$. It is seen from (5) that a GB has a Gaussian amplitude transverse to the beam axis (z axis) for a given z, with the $1/e$ boundaries occurring at $x = \pm w(z)$. The beam waist is defined as the $1/e$ width or "spot size" of the beam at its narrowest point, i.e., at $z = 0$, and is given by

$$w_z = 2w(0) = 2\sqrt{\frac{2b}{k}}. \tag{8}$$

It is noted that this 2-D GB basis function can also be derived as the paraxial solution of the wave equation for a line source located at the complex location $(x', z') = (0, -jb)$ [9]. Therefore, a GB independently satisfies Maxwell's equations in the paraxial region, and for a well-focused beam the fields are confined primarily to this region.

The far field form of the GB basis function is obtained by making the transformation to polar coordinates

$$x = \rho \sin\theta, \; z = \rho \cos\theta \tag{9}$$

and letting ρ approach infinity. In the paraxial region (i.e., for small values of θ) this yields

$$B(\rho, \theta; b) = \sqrt{jb}\frac{e^{-jk\rho}}{\sqrt{\rho}} e^{-\frac{1}{2}kb\theta^2}. \tag{10}$$

Notice from this result that in the far field a GB is also Gaussian in angle (in the paraxial region), and that it propagates with a cylindrical spread factor. The angular beam width is defined as the $1/e$ beam width of the far field basis function and is given by

$$w_\theta = 2\sqrt{\frac{2}{kb}}. \tag{11}$$

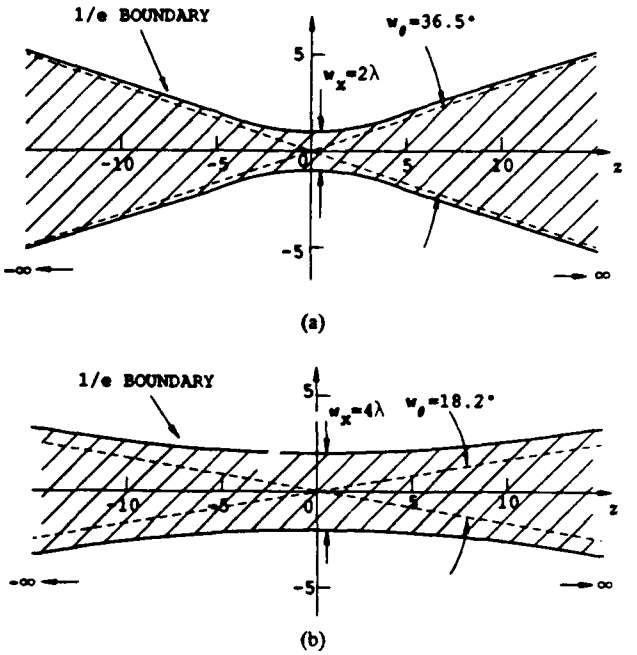


Fig. 4. Two GB's shaded within their $1/e$ amplitude boundaries. (a) $b = \pi\lambda$, narrow waist, wide angular beam width (divergent). (b) $b = 4\pi\lambda$, wide waist, narrow angular beam width (well focussed).

This quantity is sometimes referred to as the beam divergence angle, as in [22].

Figure 4 shows two typical GB's with different beam parameters illustrating the effect of b on the beam waist and the angular beam width. The GB depicted in Fig. 4(a) is not well focused as compared to the one in Fig. 4(b) which is relatively well focused. The reader is referred to the introduction for the physical reasons that influence the choice of initial launch parameters for the well-focused beams used in the axial tracking approximation, and the various trade-offs that must be considered when using this approximation.

The beam expansion coefficients $A_{mn}(\theta_i)$ in (3) are found by discretizing the aperture into subapertures of width L_x which are centered at the array points mL_x, and requiring that the set of angularly shifted (i.e., rotated) GB's launched radially from the center of each subaperture reproduce the far fields of that subaperture. The expansion coefficients of the GB's for a given subaperture, as well as the constant interbeam spacing $\Delta\theta$, are then found by point matching the rotated GB's to the far field radiation pattern of the subaperture from which the GB's are launched so that the subaperture pattern can be adequately reproduced by the beam expansion. Such a procedure for finding the expansion coefficients from the far field information appears to be far simpler than finding them directly from the field distribution in the aperture because the rotated beams of a particular subaperture are distinctly separated in angle in the far field, whereas they are superimposed on top of each other in the spatial subaperture interval.

The field radiated in the far zone of the mth subaperture is expressed as a sum of the angularly rotated GB's of the mth subaperture using the far field beam function of (10), i.e.,

$$U_m(\rho,\theta)=\sum_n A_{mn}(\theta_i)\sqrt{jb}\frac{e^{-jk\rho}}{\sqrt{\rho}}e^{-\frac{1}{2}kb(\theta-\theta_n)^2} \quad (12)$$

where (ρ,θ) are the polar coordinates of the mth subaperture, with ρ being the distance from the phase center of the mth subaperture to the far field and θ measured from the $+\hat{z}$ axis. $\theta_n = n\Delta\theta$ is the direction in which the nth GB is launched. The expansion of (12) is truncated at some angle $\pm\theta_{max}$ which is close to but less than 90°, so that $-\theta_{max} \le \theta_n \le \theta_{max}$. The field radiated by the mth subaperture which is illuminated by a plane wave incident from the direction θ_i with amplitude U^i is found via the Kirchhoff approximation as follows:

$$U_m(\rho,\theta)=f_m(\theta,\theta_i)\frac{e^{-jk\rho}}{\sqrt{\rho}}$$

$$f_m(\theta,\theta_i)=U^i\sqrt{\frac{jk}{8\pi}}e^{-jkmL_x\sin\theta_i}L_x(\cos\theta+\cos\theta_i)\cdot\operatorname{sinc}\left[\frac{1}{2}kL_x(\sin\theta-\sin\theta_i)\right] \quad (13)$$

where $\operatorname{sinc}(u)=(\sin u)/u$. A_{mn}, L_x, b, and $\Delta\theta$ should be chosen so that the expansion of (12) adequately reproduces (13) using beams which satisfy the physical constraints of the problem considered here. While there is considerable freedom in making these choices, the following procedure has been found through experience to yield accurate results with good efficiency.

For the expansion of (12) to adequately reproduce (13) via a point matching or "sampling" procedure, the function $f(\theta,\theta_i)$ of (13) is assumed to be slowly varying with respect to $\Delta\theta$ and that in the direction θ_n, only the nth and $(n\pm 1)$th beams of the mth subaperture are significant; all other beams of the mth subaperture are negligible because of their Gaussian amplitude taper. Then it is reasonable to make $A_{mn}(\theta_i)$ proportional to $f_m(\theta_n,\theta_i)$. It has been found using these assumptions, and through numerical experiment, that the following relationships used in the expansion of (12) adequately reproduce (13):

$$\Delta\theta=\sqrt{\frac{2\log 4}{kb}} \quad (14)$$

$$A_{mn}(\theta_i)=\frac{1}{\log 4\sqrt{jb}}f_m(\theta_n,\theta_i) \quad (15)$$

$$\Delta\theta\le\frac{\lambda}{3.5L_x}. \quad (16)$$

The inequality in (16) guarantees that at least 3.5 GB's will make up each lobe of the far field pattern $f_m(\theta,\theta_i)$.

It still remains to find the beam parameter b, which then determines $\Delta\theta$, $A_{mn}(\theta_i)$ and the subaperture size L_x via (14), (15), and (16), respectively. In practice, the beam waist w_x is chosen first to produce GB's which best fit the particular geometry under consideration. The initial beam waist size then determines b through (8). In choosing w_x it is best to require that the initial beam waist be less than one-half the width d of the aperture. However, to have beams which are initially relatively well focused, it is necessary to require that the initial beam waist is larger than approximately four wavelengths. The limits on the beam waist are then expressed by

$$4\lambda < w_x < \frac{1}{2}d. \quad (17)$$

If the inequality of (17) allows, a good choice of w_x is between one-third and one-half the aperture width d.

If the far field expansion of (12) adequately reproduces the far field patterns of each subaperture given by (13), then the superposition of all the GB's of all the subapertures should adequately reconstruct the incident plane wave field in the aperture. Figure 5 compares the incident plane wave field in the aperture of a cavity with its GB expansion of (3), using the guidelines of (14) thru (17). As the figure shows, the GB's reproduce the aperture field quite well, considering that these fields are abruptly truncated by the Kirchhoff approximation.

III. Gaussian Beam Axial Tracking Approximation

The GB's in the expansion of (3) are tracked within the cavity to the termination via multiple reflections from the interior walls using an axial approximation for the reflection of a beam from a curved surface. In this approximation it is assumed that a GB incident on a curved surface gives rise to a new reflected GB whose beam axis intersects the surface at the same point as the incident beam axis. The parameters of the reflected beam are related to those of the *incident* beam by matching of the fields on the reflecting surface in the vicinity of the point where the beam axes intersect. Referring to Fig. 6, the z' axis denotes the axis of the incident beam and the z axis denotes the axis of the reflected beam; let the incident GB field U^i in Fig. 6 be represented as

$$U^i(x',z')=U_o'B(x',z';b') \quad (18)$$

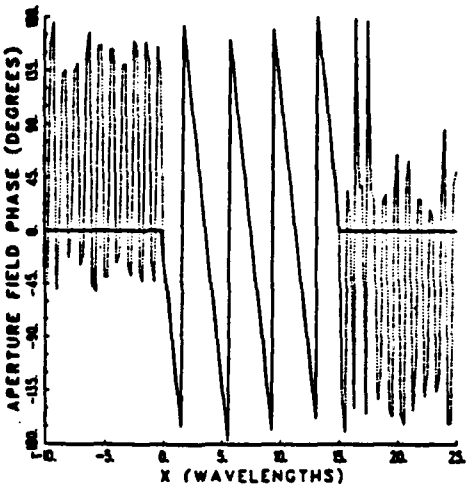
and the reflected GB field U^r be represented as

$$U^r(x,z)=U_oB(x,z;b) \quad (19)$$

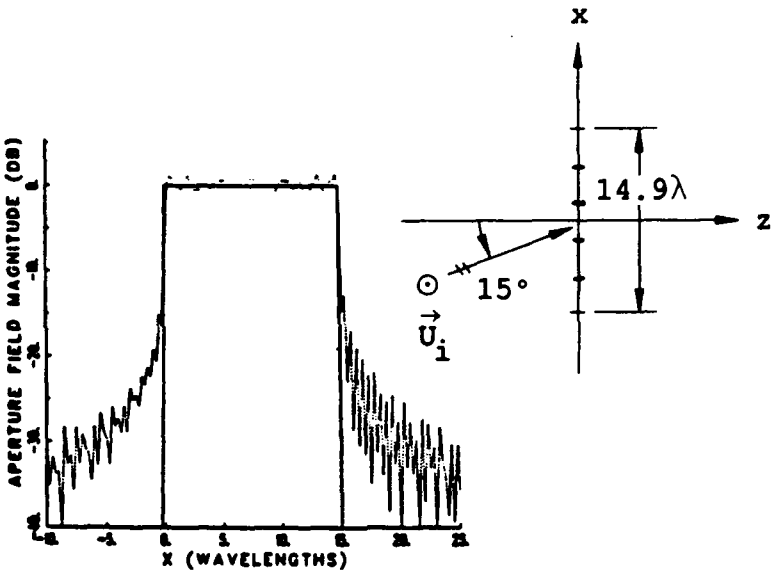
where U_o' and U_o are constants and $\mathcal{B}$ is the GB basis function of (5) (note that the primed quantities refer to the incident beam). In Fig. 6, ρ_c' and ρ_c are the distances from the waists of the incident and reflected GB's, respectively, to the point where the beam axes intersect the surface; either quantity may be positive or negative depending on whether the beam is diverging or converging, respectively, at the reflecting surface. R_c is the radius of curvature of the surface, and θ_i and θ_r are the angles the incident and reflected beam axes, respectively, measured from the surface normal; all of these quantities are defined at the point of reflection of the incident beam axis on the surface.

First, the incident and reflected GB's are matched exactly at the point where the beam axes intersect the surface. This yields the relationship

$$U_o=\Gamma(\theta_i)U_o'\sqrt{\frac{b'(\rho_c+jb)}{b(\rho_c'+jb')}}e^{-jk(\rho_c'-\rho_c)} \quad (20)$$



(a)



(b)

Fig. 5. GB reconstruction of the aperture field due to plane wave incidence. $w_x = 7.5\lambda$, 5 subapertures, $L_x = 2.98\lambda, \Delta\theta = 6.0°$, $\theta_{max} = 60°$. —— incident plane wave field, - - - - GB recontruction. (a) Phase of the aperture field. (b) Magnitude of the aperture field.

where $\Gamma(\theta_i)$ is the appropriate Fresnel plane wave reflection coefficient of the surface, evaluated for the axial incident angle θ_i. Next, the exponential portions of both the incident and reflected beam basis functions are expanded in Taylor series as a function of the distance along the surface from the point where the beam axes intersect. Matching the linear and quadratic terms of these two expansions yields

$$\theta_r = \theta_i \tag{21}$$

$$\frac{1}{\rho_c + jb} = \frac{1}{\rho_c' + jb'} + \frac{2}{R_c \cos\theta_i}. \tag{22}$$

Notice that these expressions are similar to the laws of geometrical optics governing ray reflection from a curved surface, but with complex caustic locations. This is not surprising considering that the incident 2-D GB can be thought of as the paraxial field of a line source moved into complex space by an amount b'. Similar expressions can be derived for the 3-D case which are also closely analogous to the laws of GO.

It is noted that matching higher order terms in the Taylor expansions gives results inconsistent with our prior assumption that an incident GB gives rise to a reflected GB at the point where the axes intersect the surface. This is because realistically, an incident GB gives rise to a nonsymmetric reflected field which is only approximately a GB [14]. However, for cases when the incident beam illuminates an area on the surface which is small with respect to the radius of curvature at that point, our approximation is reasonable. Therefore, the above expressions are valid as long as R_c is much larger than the width of the incident beam where it intersects the curved surface and as long as the beam does not come close to grazing the surface (i.e., for θ_i not close to 90°).

In general, the farther a GB propagates inside a cavity the

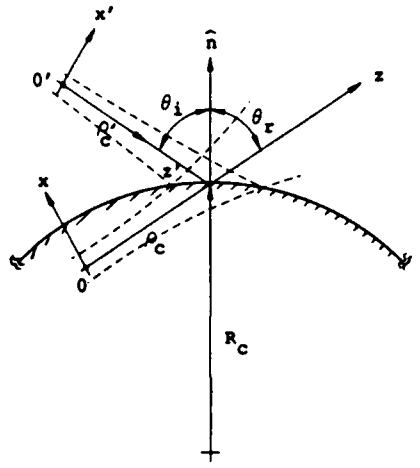


Fig. 6. Axial reflection of a GB from a curved surface.

more it diverges until it becomes too wide to fit inside the cavity, or too wide to satisfy the beam axial approximation. This tends to restrict the overall length of cavities for which the GB shooting method can be applied within the axial tracking approximation. The method works best for electrically large cavities because the beams stay well focused. An approximate limit on the length L of a cavity of width d for which the GB shooting method with axial tracking can be applied has been obtained through analysis and numerical experimentation as

$$L \leq 0.2d^2/\lambda. \tag{23}$$

IV. Termination Reciprocity Integral Formulation

Once the fields inside a duct have been tracked to the vicinity of the termination via GB shooting or some other appropriate method, the termination reciprocity integral can be used to find the cavity scattered fields. This integral is derived in [21] and [2] and will be described briefly in this section.

Consider the 3-D cavity geometry with an arbitrary termination as shown in Fig. 7. This cavity is illuminated by two infinitesimal dipole sources $d\vec{p}_a$ and $d\vec{p}_b$ located at Q_a and Q_b, respectively. Let the total fields due to $d\vec{p}_a$ and $d\vec{p}_b$ radiating in the presence of the cavity and its termination be denoted by $\vec{E}_a$, $\vec{H}_a$ and $\vec{E}_b$, $\vec{H}_b$, respectively. These fields can be decomposed as

$$\vec{E}_a = \vec{E}_a^i + \vec{E}_a^s \,;\vec{H}_a = \vec{H}_a^i + \vec{H}_a^s \tag{24}$$

$$\vec{E}_b = \vec{E}_b^i + \vec{E}_b^s \,;\vec{H}_b = \vec{H}_b^i + \vec{H}_b^s \tag{25}$$

where $(\vec{E}_a^i, \vec{H}_a^i)$ and $(\vec{E}_b^i, \vec{H}_b^i)$ are the fields due to $d\vec{p}_a$ and $d\vec{p}_b$, respectively, in the presence of the cavity but with the termination removed (or absent), so $(\vec{E}_a^s, \vec{H}_a^s)$ and $(\vec{E}_b^s, \vec{H}_b^s)$ are the fields due to $d\vec{p}_a$ and $d\vec{p}_b$, respectively, which are scattered by the termination. The termination reciprocity integral is then stated as follows:

$$\vec{E}_b^s(Q_a) \cdot d\vec{p}_a = \int_{S_t} \left(\vec{E}_b^s \times \vec{H}_a^i - \vec{E}_a^i \times \vec{H}_b^s \right) \cdot \hat{n} dS \tag{26}$$

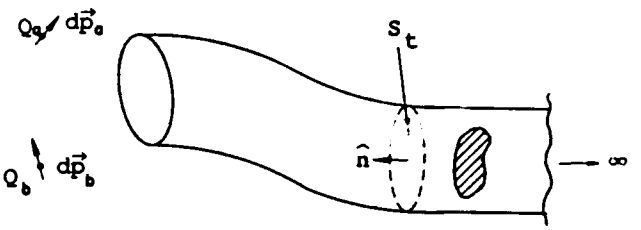


Fig. 7. 3-D cavity illuminated by two infinitesimal dipole sources.

where S_t is a conveniently located cross section inside the cavity near the termination, as shown in Fig. 7, and $\hat{n}$ is the unit surface normal of S_t pointing away from the termination. To use (26), it is assumed that $\vec{E}_a^i$ and $\vec{H}_a^i$ are known by tracking the fields from $d\vec{p}_a$ through the interior of the cavity to S_t in the absence of the termination, and that $\vec{E}_b^s$ and $\vec{H}_b^s$ are known by tracking the fields $\vec{E}_b^i, \vec{H}_b^i$ from $d\vec{p}_b$ through the interior of the cavity to the termination and tracking them back to S_t after being scattered by the termination.

Finding the interior scattered field at S_t can be very difficult for an arbitrary termination; some procedures for accomplishing this are currently under study and will be reported separately in the future. In the numerical results presented in this paper, a simple planar perfectly conducting termination is employed. It is noted that (26) intrinsically includes all the multiple interactions between the termination and the open end of the cavity. However, these multiple wave interactions are ignored for the reasons given earlier in the introduction.

For the case of plane wave illumination, far field scattering, and a planar termination of width l which is normal to the z axis, it is straightforward to show that (26) reduces to the following expression for the 2-D case:

$$U_b^s(Q_a) = -\frac{1}{U^i\sqrt{j8\pi k}} \frac{e^{-jk\rho}}{\sqrt{\rho}} \int_l \left(U_b^s \frac{\partial U_a^i}{\partial z} - U_a^i \frac{\partial U_b^s}{\partial z} \right) dx \tag{27}$$

where $U_b^s(Q_a)$ is the bistatic scattered field in the direction of Q_a at a far field distance of ρ from the open end of the cavity, due to a plane wave of amplitude U^i incident from the direction of Q_b. The integration in (27) is over the termination boundary. U_a^i in the integrand of (27) is found by launching GB's which are excited by the field incident at the open end from Q_a, and tracking them through the cavity interior to S_t. The GB launching expansion is given in (3). Similarly, U_b^s in the integrand of (27) is found by launching the GB's excited by the field incident at the open end from Q_b, and tracking them through the cavity to the termination and reflecting them back to the termination boundary S_t. For the case when S_t lies directly on a planar termination, U_b^s can be easily found from U_b^i by using the boundary conditions on the termination surface.

V. Results and Conclusions

The numerical results presented here are for the case of backscatter from 2-D open waveguide cavity geometries which are made up of straight and annular perfectly conducting waveguide sections and are terminated by a planar perfect conductor. An independent hybrid modal reference solution as well as a ray shooting (SBR) solution [1], [2] is available for these types of geometries. The numerical results include the external scattering by the edges at the open end along with the internal cavity scattering. No other external scattering effects are considered. Furthermore, the multiple interactions between the open end and the termination are neglected for reasons mentioned earlier. The numerical results are presented as echo width in decibels (dB) relative to a wavelength versus incidence angle. The 2-D echo width is defined by

$$\sigma = \lim_{\rho \to \infty} 2\pi\rho \frac{|U_b^s(Q_a)|^2}{|U^i|^2} \tag{28}$$

where $U_b^s(Q_a)$ is calculated from (27) for the case of a perfectly conducting planar termination coinciding with the plane of integration, and for scattering in the direction of the incident plane wave (backscatter). For the results obtained here via GB shooting, the initial beam waist w_x is specified first to launch GB's which fit well within the open end of the cavity, and the initial beam parameter b is found from (8) which then determines $\Delta\theta$ from (14). The subaperture size L_x is chosen to be as large as possible under the constraint of (16).

Figure 8 compares the backscatter patterns of a straight cavity calculated using three different methods. The GB result compares quite well with the hybrid modal reference solution, whereas the ray shooting (SBR) result predicts the overall trends in the pattern but misses some of the details. The inaccuracy of the SBR result is due to exclusion of the fields coupled into the cavity via diffraction of the incident plane wave by the leading edges; on the other hand, the subaperture expansion used in the GB shooting method implicitly includes the effects of these fields to within the Kirchhoff approximation. A total of 125 GB's were tracked just once, independent of incidence angle, to obtain the results of Fig. 8. In contrast, the SBR method required one to track 200 rays for each incidence angle at one degree increments, so a total of 12 000 rays were tracked! Of course, the number of GO rays tracked would be reduced if the backscatter was calculated at larger increments of the incidence angle, thus sacrificing some continuity in the pattern.

Figure 9 shows the backscatter patterns of a relatively short double-bend S-shaped cavity. The two GB shooting solutions, which use different initial beam parameters, agree well with each other as well as with the reference solution. It should be mentioned here that in realistic scattered field calculations, a result which is accurate to within 3–5 dB is usually considered a good approximation for most practical purposes, particularly in the vicinity of the peaks of the scattering pattern. The results of Fig. 9 support the capabilities of the GB method for practical applications.

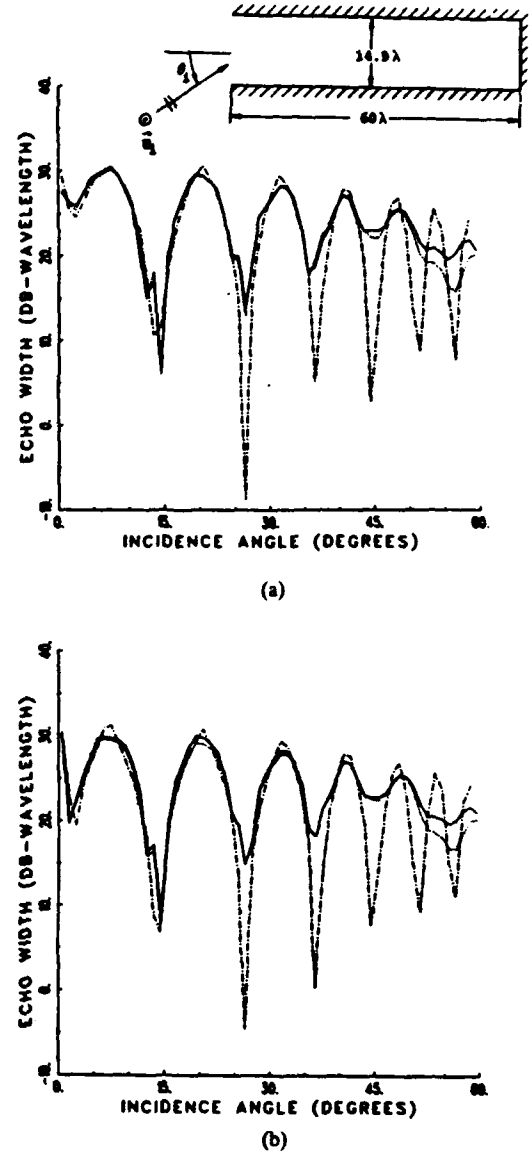
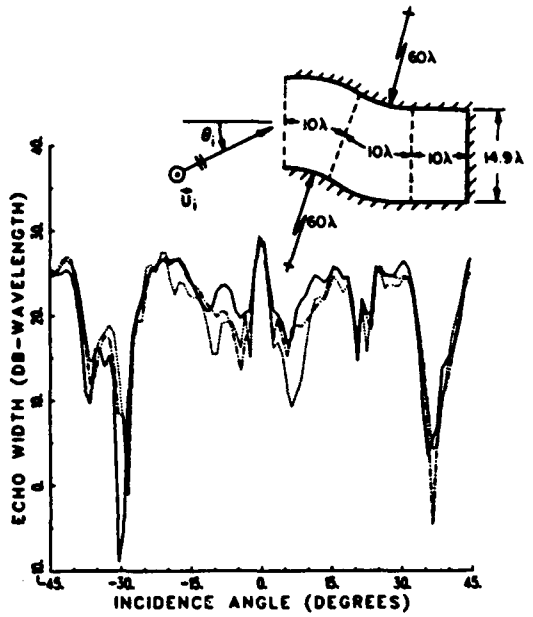
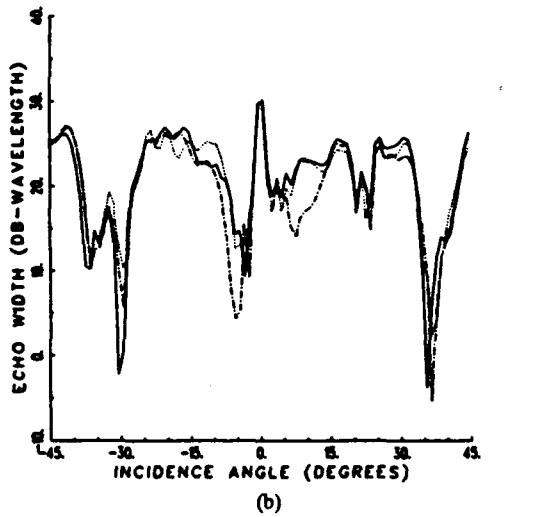


Fig. 8. Backscatter patterns of a parallel-plate cavity. ——— hybrid modal reference solution, - - - - - GB shooting method, $w_x = 7.5\lambda$, $L_x = 2.98$, $\theta_{max} = 75°$, — - — SBR ray shooting method, 200 ray-tubes/angle. (a) TM_y polarization. (b) TE_y polarization.

The reason that the GB results in Fig. 8 are somewhat better than those in Fig. 9 is because the curved walls of the double-bend geometry in the latter case cause the beams to diverge and distort after reflection, introducing small errors in the axial beam tracking approximation. The farther a beam propagates inside a cavity with curved walls, the more these errors accumulate. This is what limits the length to approximately $.2d^2/\lambda$ as in (23). For an aperture width of 14.9λ, the limit on the length is approximately 45λ. Figure 10 shows the backscatter patterns of a double-bend cavity which is longer than this limit. The agreement is still good, but this may not always be the case as the pattern of Fig. 11(a) pertains to a double-bend cavity of the same axial length but with a somewhat different geometry. However, as Fig. 11(b) shows, the agreement improves when the frequency is doubled. It is noted that the straight cavities

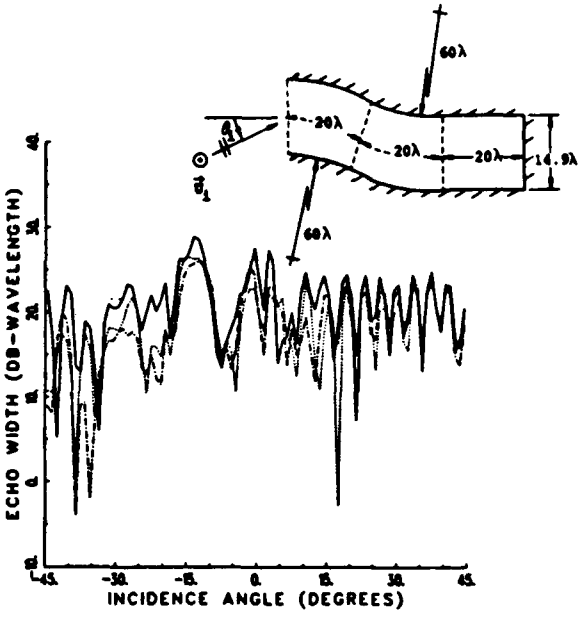


(a)

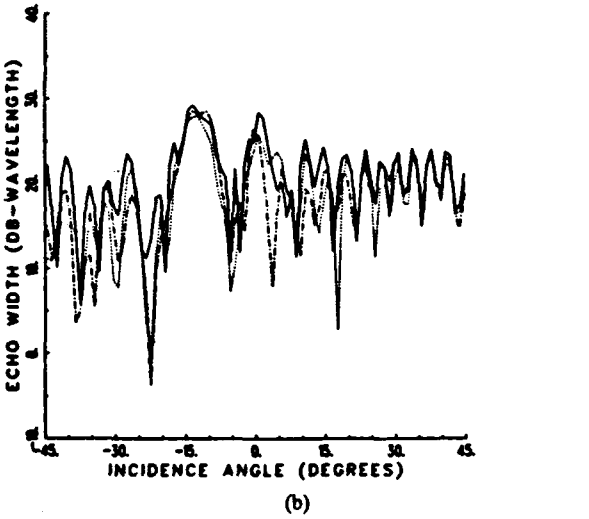


(b)

Fig. 9. Backscatter patterns of a short double-bend S-shaped cavity. ——— hybrid modal reference solution, - - - - - GB shooting method, $w_x = 7.5\lambda$, $L_x = 2.98$, $\theta_{max} = 60°$, — - — GB shooting method, $w_x = 5.0\lambda$, $L_x = 1.66$, $\theta_{max} = 60°$. (a)TM_y polarization. (b) TE_y polarization.



(a)



(b)

Fig. 10. Backscatter patterns of a long double-bend S-shaped cavity. —— hybrid modal reference solution, - - - - - GB shooting method, $w_x = 7.5\lambda$, $L_x = 2.98$, $\theta_{max} = 60°$, — · — GB shooting method, $w_x = 5.0\lambda$, $L_x = 1.66$, $\theta_{max} = 60°$. (a) TM_y polarization. (b) TE_y polarization.

which can be analyzed using the GB shooting method can be longer than the limit given in (23), as in the case of Fig. 8, because the flat walls do not cause errors to accumulate.

It has been shown that the GB shooting method can be useful in predicting the electromagnetic coupling into and scattering by large open-ended nonuniform waveguide cavities with smooth, slowly varying interior walls for which the axial beam tracking approximation employed here remains applicable. This method therefore works well for nonuniform cavities which are wider than approximately eight wavelengths and have a length to width ratio of less than approximately one-fifth of the width in wavelengths. It is noted of course, that the length of the cavity could be increased further without difficulty if the axial tracking approximation is replaced by the more accurate procedure for tracking beams via complex rays, but that would be done with much less efficiency. Finally, the termination reciprocity integral employed here allows the external scattered fields to be found in terms of the fields tracked one-way from the open end of the cavity to the interior termination, and allows the cavities to have arbitrary terminations as long as the local internal reflection characteristics of such terminations are known.

REFERENCES

[1] P. H. Pathak and R. J. Burkholder, "Modal, ray, and beam techniques for analyzing the EM scattering by open-ended waveguide cavities," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 635–647, May 1989.

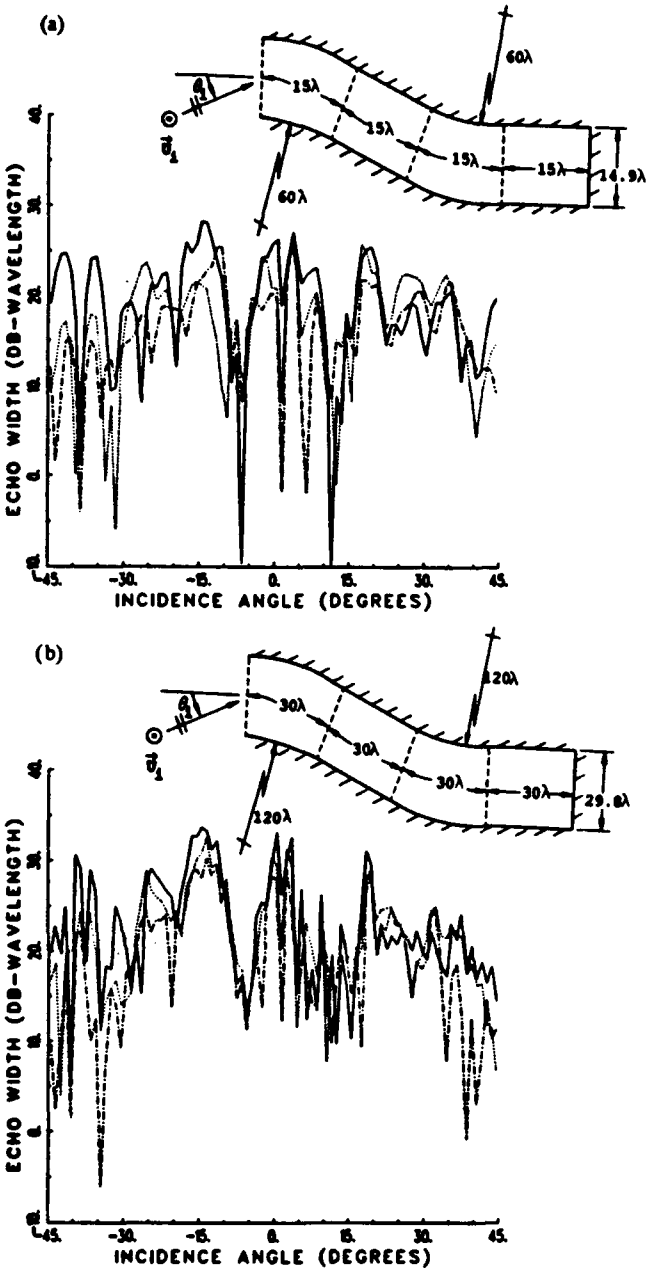


Fig. 11. Effect of doubling the frequency on the backscatter pattern of a long double-bend S-shaped cavity, TM_y polarization. ——— hybrid modal reference solution. - - - - - GB shooting method, $w_x = 7.5\lambda$, $\theta_max = 60°$, — - — GB shooting method, $w_x = 5.0\lambda$, $\theta_max = 60°$, (a) 14.9λ aperture. (b) 29.8λ aperture.

[2] R. J. Burkholder, "High-frequency asymptotic methods for analyzing the EM scattering by open-ended waveguide cavities," Ph. D. dissertation, The Ohio State University, Columbus, June 1989.

[3] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, pp. 116–130, 1962.

[4] R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, vol. 62, pp. 1448–1461, Nov. 1974.

[5] P. H. Pathak, "Techniques for high frequency problems," ch. 4 in *Antenna Handbook—Theory, Applications, and Design*, Y. T. Lo and S. W. Lee, Eds. New York: Van Nostrand Reinhold, 1988.

[6] P. Ya. Ufimtsev, "Method of edge waves in the physical theory of diffraction," from the Russian "Method Krayevykh voin v fizicheskoy teoril difraktsii," *Izd-Vo Sov. Radio*, pp. 1–243, 1962, translation prepared by the U.S. Air Force Foreign Technology Division, Wright-Patterson Air Force Base, Ohio; released for public distribution Sept. 7, 1971.

[7] S. W. Lee, "Comparison of Uniform Asymptotic Theory and Ufimtsev's Theory of EM edge diffraction," *IEEE Trans. Antennas Propagat.*, vol. AP-25, pp. 162–170, Mar. 1977.

[8] H. Ling, R. Chou, and S. W. Lee, "Shooting and bouncing rays: Calculating RCS of an arbitrary cavity," IEEE Trans. Antennas Propagat., vol. 37, pp. 194–205, Feb. 1989.

[9] G. A. Deschamps, "The Gaussian beam as a bundle of complex rays," Electron. Lett., vol. 7, no. 23, pp. 684–685, Nov. 18, 1971.

[10] G. A. Deschamps, "Ray techniques in electromagnetics," *Proc. IEEE*, vol. 60, Sept. 1972.

[11] J. W. Ra, H. L. Bertoni, and L. B. Felsen, "Reflection and transmission of beams at a dielectric interface," *SIAM J. Appl. Math.*, vol. 24, no. 3, pp. 396–413, May 1973.

[12] S. Y Shin and L. B. Felsen, "Multiply reflected Gaussian beams in a circular cross section," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26, pp. 845–851, Nov. 1978.

[13] F. J. V. Hasselmann and L. B. Felsen, "Asymptotic analysis of parabolic reflector antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 677–685, July 1982.

[14] Y. Z. Ruan and L. B. Felsen, "Reflection and transmission of beams at a curved interface," *J. Opt. Soc. Amer.*, vol. 3, no. 4, pp. 566–578, Apr. 1986.

[15] J. J. Maciel and L. B. Felsen, "Gaussian beam analysis of propagation from an extended plane aperture distribution through plane and curved dielectric layers," parts I and II, *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 1607–1624, Oct. 1990.

[16] H. Shirai and L. B. Felsen, "Rays, modes and beams for plane wave coupling into a wide open-ended parallel-plane waveguide," *Wave Motion*, vol. 9, pp. 301–317, 1987.

[17] H. Shirai and L. B. Felsen, "Rays and modes for plane wave coupling into a large open-ended circular waveguide," *Wave Motion*, vol. 9, pp. 461–482, 1987.

[18] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.* (London) vol. 93III, pp. 429–457, 1946.

[19] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. IEEE*, vol. 68, pp. 538–539, 1980.

[20] J. J. Maciel and L. B. Felsen, "Systematic study of fields due to extended apertures by Gaussian beam discretization," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 884–892, July 1989.

[21] P. H. Pathak and R. J. Burkholder, *Internal memorandum* on the use of a generalized reciprocity theorem to deal with the scattering by an arbitrary termination within an otherwise open-ended cavity, 1988.

[22] J. T. Verdeyen, in *Laser Electronics*. Englewood Cliffs, NJ: Prentice-Hall, 1981, ch. 3.

Robert J. Burkholder (Member, IEEE) received the B.S. , M.S., and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1984, 1985, and 1989, respectively.

As an undergraduate he worked as a co-op student for General Electric in East Cleveland, OH. He is currently with The Ohio State University as a Post-doctoral Research Associate. In his dissertation, he contributed to the development of modal, ray, and beam approaches for analyzing the electromagnetic (EM) fields scattered from and coupled into electrically large, nonuniform open-ended waveguide cavities. His research interests are in the areas of high frequency, modal and hybrid techniques for solving EM antenna and scattering problems. Currently, he is working on the analysis of the RCS of open cavities and partially coated conducting objects.

Prabhakar H. Pathak (Fellow, IEEE) received the B.Sc. degree in physics from the University of Bombay, India, in 1962, the B.S. degree in electrical engineering from the Louisiana State University, Baton Rouge, in 1965, and the M.S. and Ph.D degrees in electrical engineering from The Ohio State University, Columbus, in 1970 and 1973, respectively.

From 1965 to 1966 he was an Instructor in the Department of Electrical Engineering at the University of Mississippi, Oxford. During the summer of 1966, he was an Electronics Engineer with the Boeing Company in Renton, WA. Since 1968 he has been with the ElectroScience Laboratory, The Ohio State University, initially as a Graduate Research Associate, and later as a member of the professional research staff after graduating in 1973. Since 1982, he has also been on the Faculty of the Department of Electrical Engineering, The Ohio State University, teaching courses in electromagnetics and in linear systems. His research work has dealt primarily with the development of uniform ray solutions which improve and extend the geometrical theory of diffraction for analyzing EM antenna and scattering problems of engineering interest. Currently, his research activities continue to be in the areas of mathematical methods, EM antenna and scattering problems, and uniform ray techniques. He has been an invited lecturer at many short courses on GTD and high frequency techniques both in the U.S. and abroad. He has been a contributor to chapters on high frequency techniques for five books.

Angle Estimation Using a Polarization Sensitive Array

Jian Li and R. T. Compton, Jr.

Abstract—**In a previous paper, the authors described how the ESPRIT algorithm may be used to estimate both the arrival directions and the polarizations of incoming plane waves with a uniform linear array of crossed dipoles. An alternative approach is described here that may be used with the same array to estimate arrival angles only. This new approach has the advantage that it requires far fewer computations. The performance of this approach is compared with that of the original approach and with that of estimators using conventional ESPRIT arrays.**

I. Introduction

In a previous paper [1], the authors described how the ESPRIT algorithm [2] can be used to estimate both the arrival directions and the polarizations of incoming plane waves with a uniform linear array of crossed dipoles. The ESPRIT algorithm exploits the invariance properties of such an array so that both angle and polarization estimates can be obtained.

In some applications, however, one may be interested in estimating only the signal directions, but not the polarizations. The purpose of this communication is to show how the approach in [1] can be simplified when polarization estimates are not needed. The simplification consists of averaging the two covariance matrices associated with two orthogonal polarizations. We compare the performance of this approach with that of the earlier approach in [1] and also with that of a conventional ESPRIT estimator that does not take polarization into account.

Manuscript received December 17, 1990; revised June 24, 1991.

This work was supported in part by the Joint Services Electronics Program under Contract N00014-89-J-1007 with The Ohio State University Research Foundation.

J. Li was with the Department of Electrical Engineering, The Ohio State University, Columbus, OH. She is now with the Department of Electrical Engineering, University of Kentucky, Lexington, KY 40506.

R. T. Compton, Jr. is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210.

IEEE Log Number 9102628.

II. Problem Formulation

Consider a $2L$-element array consisting of L pairs of crossed dipoles, as shown in Fig. 1. Each dipole in the array is a short dipole, so the output voltage from each dipole is proportional to the electric field component along that dipole. The signal from each dipole is to be processed separately by the estimator. The lth dipole pair, $l = 1, 2, \cdots, L$, has its center on the y-axis at $y = (l-1)\delta$. The distance δ between two adjacent dipole pairs is assumed to be a half-wavelength to avoid angle ambiguity problems.

Suppose K (with $K < L$) continuous wave (CW) signals impinge on the array from angular directions θ_k in the yz-plane, where θ denotes the polar angle in the yz-plane as shown in Fig. 1 and $k = 1, 2, \cdots, K$. The electric field for signal k is given by

$$\mathbf{E}_k = \left(-E_{x_k}\mathbf{e}_x + E_{\theta_k}\mathbf{e}_{\theta_k}\right)e^{j(\omega t - \mathbf{k}_k\cdot\mathbf{r} + \psi_k)}, \tag{1}$$

where E_{x_k} and E_{θ_k} are the electric field components in the $-x$ and θ_k directions, respectively, and $\mathbf{e}_x$ and $\mathbf{e}_{\theta_k}$ denote unit vectors in the x- and θ_k-directions. $\mathbf{k}_k$ is the propagation vector for signal k, given by $\mathbf{k}_k = -k_0\mathbf{e}_{r_k}$, where k_0 is the free-space propagation constant and $\mathbf{e}_{r_k}$ is a unit vector pointing outward along the radial direction defined by angle θ_k. (In (1), the minus sign on E_{x_k} is included because $-\mathbf{e}_x, \mathbf{e}_{\theta_k}$, and $-\mathbf{e}_{r_k}$, in that order, form a right-handed coordinate system for an incoming signal.) The vector $\mathbf{r}$ in (1) is a vector from the coordinate origin to a point in space where the field is measured. Finally, ω is the frequency and ψ_k the carrier phase angle of signal k. ψ_k is assumed to be a random variable uniformly distributed on $[0, 2\pi)$, and the ψ_k for different signals are assumed statistically independent of each other.

We assume each signal has an arbitrary electromagnetic polarization [3]. To specify the signal polarizations, we let α_k and β_k denote the ellipticity angle and the orientation angle, respectively, of the polarization ellipse produced by E_{x_k} and E_{θ_k} for signal k, as shown in Fig. 2. The angle β_k is measured with respect to the $-x$-direction. To eliminate ambiguities, β_k is defined to be in the range $0 \le \beta_k < \pi$, and α_k is always in the range $-\pi/4 \le \alpha_k \le \pi/4$ [3].

0018-926X/91$01.00

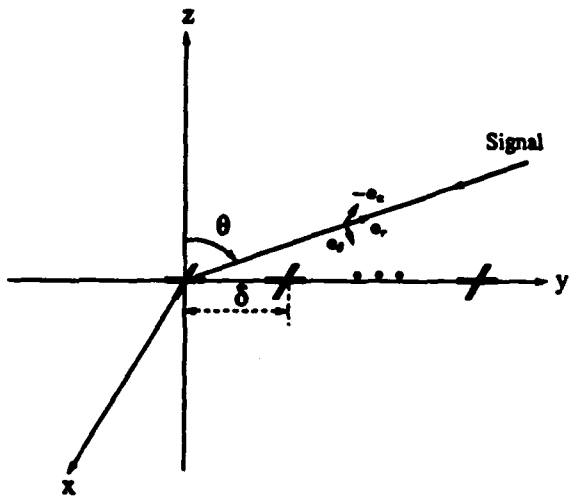


Fig. 1. A uniform linear array of crossed dipoles.

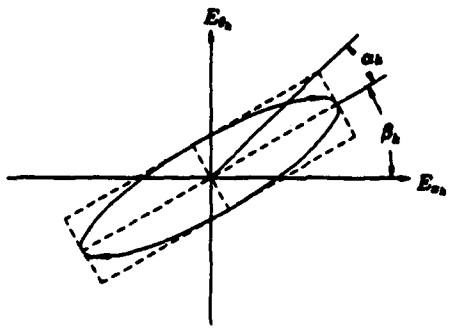


Fig. 2. Polarization ellipse.

In terms of α_k and β_k, the field components E_{x_k} and E_{θ_k} are given by

$$E_{x_k} = E_k \cos \gamma_k \tag{2}$$

$$E_{\theta_k} = E_k \sin \gamma_k e^{j\eta_k} \tag{3}$$

where E_k is the amplitude and γ_k and η_k are related to α_k and β_k by

$$\cos 2\gamma_k = \cos 2\alpha_k \cos 2\beta_k, \tag{4}$$

$$\tan \eta_k = \tan 2\alpha_k \csc 2\beta_k. \tag{5}$$

(A proof of the relations in (3)–(5) can be found in [4].) By combining (1)–(5), the electric field for signal k may be written

$$\mathbf{E}_k = E_k\left[-\cos \gamma_k \mathbf{e}_x + \sin \gamma_k e^{j\eta_k}\mathbf{e}_{\theta_k}\right] e^{j(\omega t - \mathbf{k}_k \cdot \mathbf{r} + \psi_k)} \tag{6}$$

$$= s_k(t)\left[-\cos \gamma_k \mathbf{e}_x + \sin \gamma_k e^{j\eta_k}\mathbf{e}_{\theta_k}\right] e^{-j\mathbf{k}_k \cdot \mathbf{r}}, \tag{7}$$

where $s_k(t)$ is the kth incident signal,

$$s_k(t) = E_k e^{j(\omega t + \psi_k)}. \tag{8}$$

For the lth dipole pair, let $x_l(t)$ be the signal received from the x-axis dipole, and $y_l(t)$ the signal received from the y-axis dipole. Then the total signal vector $\mathbf{z}_l(t)$ received by the lth dipole pair is

$$\mathbf{z}_l(t) = \left[x_l(t) \quad y_l(t)\right]^T, \tag{9}$$

where $(\cdot)^T$ denotes the transpose. Assume that, in addition to the received signals, $\mathbf{z}_l(t)$ contains a thermal noise voltage vector $\mathbf{n}_l(t) = [n_{x_l}(t) \quad n_{y_l}(t)]^T$. The $\mathbf{n}_l(t)$ are assumed to be zero mean, complex Gaussian processes statistically independent of each other, with covariance $\sigma^2\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. Then $\mathbf{z}_l(t)$ is given by

$$\mathbf{z}_l(t) = \sum_{k=1}^{K} \mathbf{u}_k s_k(t) q_k^{l-1} + \mathbf{n}_l(t), \qquad l = 1,2,\cdots,L \tag{10}$$

where

$$\mathbf{u}_k = \begin{bmatrix} -\cos \gamma_k \\ \sin \gamma_k \cos \theta_k e^{j\eta_k} \end{bmatrix} \tag{11}$$

and

$$q_k = e^{j\frac{2\pi\delta}{\lambda}\sin\theta_k} \tag{12}$$

where λ is the wavelength.

Let $\mathbf{z}(t)$, $\mathbf{s}(t)$, and $\mathbf{n}(t)$ be column vectors containing the received signals, incident signals, and noise, respectively, i.e.,

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{z}_1(t) \\ \mathbf{z}_2(t) \\ \vdots \\ \mathbf{z}_L(t) \end{bmatrix}, \quad \mathbf{s}(t) = \begin{bmatrix} \mathbf{s}_1(t) \\ \mathbf{s}_2(t) \\ \vdots \\ \mathbf{s}_K(t) \end{bmatrix}, \quad \mathbf{n}(t) = \begin{bmatrix} \mathbf{n}_1(t) \\ \mathbf{n}_2(t) \\ \vdots \\ \mathbf{n}_L(t) \end{bmatrix}. \tag{13}$$

The received signal vector may then be written

$$\mathbf{z}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{14}$$

where $\mathbf{A}$ is the $2L \times K$ matrix

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \cdots \mathbf{a}_K], \tag{15}$$

whose columns are given by

$$\mathbf{a}_k = \left[\mathbf{u}_k^T \quad \mathbf{u}_k^T q_k \cdots \mathbf{u}_k^T q_k^{L-1}\right]^T. \tag{16}$$

We assume that the element signals are sampled at N distinct times t_n, $n = 1, 2, \cdots, N$. The problem of interest is to determine the angles θ_k, $k = 1, 2, \cdots, K$, from the measurements $\mathbf{z}(t_n)$, $n = 1, 2, \cdots, N$.

In a previous paper [1], the authors described how the ESPRIT algorithm may be applied to $\mathbf{z}(t_n)$, $n = 1, 2, \cdots, N$ to estimate both the arrival directions and the polarizations of the incoming plane waves. In this communication, we consider the case where we are interested in estimating the signal directions only. We describe below a simple approach that may be used when polarization estimates are not needed.

To distinguish between these two approaches, we shall refer to the earlier approach in [1] as the *full polarization method*. The simplified approach described here will be called the *angle-only method*.

III. The Simplified Approach

The central idea of the angle-only method is to consider the x-axis dipoles and the y-axis dipoles as separate subarrays. The average of the covariance matrices for the x- and y-axis subarrays is then used in the ESPRIT algorithm instead of the full covariance matrix of the crossed dipole array. The signal directions are estimated from the averaged x- and y-axis covariance matrices.

Specifically, let $\mathbf{x}(t)$ denote the column vector of signals received on the x-axis dipoles, i.e.,

$$\mathbf{x}(t) = \left[x_1(t) \quad x_2(t) \cdots x_L(t)\right]^T. \tag{17}$$

Note that $\mathbf{x}(t)$ is the subvector of $\mathbf{z}(t)$ consisting of the odd-numbered elements of $\mathbf{z}(t)$. Let $\mathbf{n}_x(t)$ be a vector containing the corresponding noise voltages,

$$\mathbf{n}_x(t) = \left[n_{x_1}(t) \quad n_{x_2}(t) \cdots n_{x_L}(t)\right]^T. \tag{18}$$

$\mathbf{n}_x(t)$ is a zero-mean complex Gaussian process with covariance $\sigma^2\mathbf{I}$. Then $\mathbf{x}(t)$ can be written

$$\mathbf{x}(t) = \bar{\mathbf{A}}_L \Phi_x s(t) + \mathbf{n}_x(t), \tag{19}$$

where $\bar{\mathbf{A}}_L$ is the direction matrix

$$\bar{\mathbf{A}}_L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ q_1 & q_2 & \cdots & q_K \\ \vdots & \vdots & \cdots & \vdots \\ q_1^{L-1} & q_2^{L-1} & \cdots & q_K^{L-1} \end{bmatrix}, \tag{20}$$

and Φ_x is defined as

$$\Phi_x = \operatorname{diag}\{-\cos\gamma_1, -\cos\gamma_2, \cdots, -\cos\gamma_K\}. \tag{21}$$

The covariance matrix of $\mathbf{x}(t)$ is given by

$$\mathbf{R}_x = E\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \bar{\mathbf{A}}_L \Phi_x \mathbf{R}_s \Phi_x^H \bar{\mathbf{A}}_L^H + \sigma^2\mathbf{I}. \tag{22}$$

Similarly, let $\mathbf{y}(t)$ denote the column vector of signals received on the y-axis dipoles, i.e.,

$$\mathbf{y}(t) = [y_1(t) \quad y_2(t) \cdots y_L(t)]^T. \tag{23}$$

$\mathbf{y}(t)$ is the subvector of $\mathbf{z}(t)$ consisting of the even-numbered elements of $\mathbf{z}(t)$. $\mathbf{y}(t)$ can be written

$$\mathbf{y}(t) = \bar{\mathbf{A}}_L \Phi_y s(t) + \mathbf{n}_y(t) \tag{24}$$

where

$$\Phi_y = \operatorname{diag}\{-\sin\gamma_1 \cos\theta_1 e^{j\eta_1}, -\sin\gamma_2 \cos\theta_2 e^{j\eta_2}, \cdots, -\sin\gamma_K \cos\theta_K e^{j\eta_K}\}, \tag{25}$$

and

$$\mathbf{n}_y(t) = [n_{y1} \quad n_{y2}(t) \cdots n_{yL}(t)]^T. \tag{26}$$

$\mathbf{n}_y(t)$ is also a zero-mean complex Gaussian process with covariance $\sigma^2\mathbf{I}$. The covariance matrix of $\mathbf{y}(t)$ is

$$\mathbf{R}_y = E\{\mathbf{y}(t)\mathbf{y}^H(t)\} = \bar{\mathbf{A}}_L \Phi_y \mathbf{R}_s \Phi_y^H \bar{\mathbf{A}}_L^H + \sigma^2\mathbf{I}. \tag{27}$$

Next let $\bar{\mathbf{R}}$ be the average of $\mathbf{R}_x$ and $\mathbf{R}_y$,

$$\bar{\mathbf{R}} = \frac{1}{2}(\mathbf{R}_x + \mathbf{R}_y). \tag{28}$$

$\bar{\mathbf{R}}$ can be written

$$\bar{\mathbf{R}} = \bar{\mathbf{A}}_L \bar{\mathbf{R}}_s \bar{\mathbf{A}}_L^H + \sigma^2\mathbf{I} \tag{29}$$

where $\bar{\mathbf{R}}_s$ is

$$\bar{\mathbf{R}}_s = \frac{1}{2}(\Phi_x \mathbf{R}_s \Phi_x^H + \Phi_y \mathbf{R}_s \Phi_y^H), \tag{30}$$

$$= \frac{1}{2}[\Phi_x \mid \Phi_y] \begin{bmatrix} \mathbf{R}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_s \end{bmatrix} [\Phi_x^H \mid \Phi_y^H]. \tag{31}$$

Since the incident signals are uncorrelated, $\mathbf{R}_s$ is nonsingular. As long as none of the incident signals produces a zero output on both the x- and y-axis dipoles at the same time, $[\Phi_x \mid \Phi_y]$ is of rank K. Then $\bar{\mathbf{R}}_s$ is also nonsingular.

To apply the ESPRIT algorithm [2], [5] to $\bar{\mathbf{R}}$, the direction matrix $\bar{\mathbf{A}}_L$ must be of full column rank. Therefore the angles of arrival must be distinct so that the columns of $\bar{\mathbf{A}}_L$ define a K-dimensional signal subspace in an L-dimensional space.

From $\bar{\mathbf{R}}$, the signal directions may be calculated as follows [2], [5]. Let $\bar{\lambda}_1 + \sigma^2 \geq \bar{\lambda}_2 + \sigma^2 \geq \cdots \geq \bar{\lambda}_K + \sigma^2 > \sigma^2 = \cdots = \sigma^2$ be the eigenvalues of $\bar{\mathbf{R}}$, and $\bar{\mathbf{e}}_1, \bar{\mathbf{e}}_2, \cdots, \bar{\mathbf{e}}_K, \bar{\mathbf{e}}_{K+1}, \cdots, \bar{\mathbf{e}}_L$ be the corresponding orthonormal eigenvectors. The columns in $\bar{\mathbf{E}}_s = [\bar{\mathbf{e}}_1 \ \bar{\mathbf{e}}_2 \cdots \bar{\mathbf{e}}_K]$ are referred to as *signal subspace eigenvectors*. They span the same signal subspace as the direction vectors in $\bar{\mathbf{A}}_L$. Let $\bar{\mathbf{E}}_{q1}$ and $\bar{\mathbf{E}}_{q2}$ be the $(L-1) \times K$ submatrices of $\bar{\mathbf{E}}_s$ consisting of the first and the last $L-1$ rows of $\bar{\mathbf{E}}_s$, respectively. Then the columns in $\bar{\mathbf{E}}_{q1}$ and $\bar{\mathbf{E}}_{q2}$ span the same subspace as the columns in $\bar{\mathbf{A}}_{L-1}$ and $\bar{\mathbf{A}}_{L-1}\Phi_q$, respectively, where

$$\Phi_q = \operatorname{diag}\{q_1, q_2, \cdots, q_K\}. \tag{32}$$

As shown in [2], the diagonal elements of Φ_q are the eigenvalues of the unique matrix $\bar{\Psi}_q$ that satisfies

$$\bar{\mathbf{E}}_{q2} = \bar{\mathbf{E}}_{q1}\bar{\Psi}_q. \tag{33}$$

From the diagonal elements of Φ_q, the signal directions θ_k can be computed from (12).

The approach described above is computationally simpler than the full polarization method in [1], because it requires the eigendecomposition of only the $L \times L$ matrix $\bar{\mathbf{R}}$ in (28). The full polarization method requires the eigendecomposition of $\mathbf{R} = E\{\mathbf{z}(t)\mathbf{z}^H(t)\}$, which is $2L \times 2L$ [1]. Hence the new approach requires approximately one eighth as many computations [6].

If the ideal array covariance matrix in (28) were known, the signal directions could be calculated exactly with ESPRIT. In practical situations, however, only a finite number of noisy measurements are taken at the dipole outputs, and the estimates of the signal directions must be made from the available measurements. Also, the number of incident signals is unknown and must be estimated. The minimum description length (MDL) criterion described by Wax and Kailath [7] can be used to estimate the number of incident signals, and the total least squares (TLS) ESPRIT algorithm [2], [8] can be used to estimate $\bar{\Psi}_q$. The steps in this process are as follows.

1) Compute

$$\hat{\bar{\mathbf{R}}} = \frac{1}{2N} = \sum_{n=1}^{N} [\mathbf{x}(t_n)\mathbf{x}^H(t_n) + \mathbf{y}(t_n)\mathbf{y}^H(t_n)], \tag{34}$$

where N denotes the number of measurements.

2) Compute the eigenvalues $\hat{\bar{\lambda}}_1 \geq \hat{\bar{\lambda}}_2 \geq \cdots \geq \hat{\bar{\lambda}}_L$ of $\hat{\bar{\mathbf{R}}}$.

3) Estimate the number of incident signals $\hat{K}$ using the MDL criterion. The MDL estimate $\hat{K}$ of the number of signals is the value of $K \in \{0, 1, \cdots, L-1\}$ that minimizes the following MDL function [7]:

$$\mathrm{MDL}(K) = -\log\left\{\frac{\Pi_{i=K+1}^{L} \hat{\bar{\lambda}}_i^{1/(L-K)}}{\frac{1}{L-K}\Sigma_{i=K+1}^{L} \hat{\bar{\lambda}}_i}\right\}^{(L-K)N} + \frac{1}{2}K(2L-K)\log N. \tag{35}$$

4) Obtain $\hat{\bar{\mathbf{E}}}_s$ whose columns are the eigenvectors of $\hat{\bar{\mathbf{R}}}$ that correspond to the $\hat{K}$ largest eigenvalues of $\hat{\bar{\mathbf{R}}}$.

5) Form $\hat{\bar{\mathbf{E}}}_{q1}$ and $\hat{\bar{\mathbf{E}}}_{q2}$ from $\hat{\bar{\mathbf{E}}}_s$ in the same way that $\bar{\mathbf{E}}_{q1}$ and $\bar{\mathbf{E}}_{q2}$ are formed from $\bar{\mathbf{E}}_s$.

6) Calculate the TLS solution $\hat{\bar{\Psi}}_q$ from $\hat{\bar{\mathbf{E}}}_{q1}$ and $\hat{\bar{\mathbf{E}}}_{q2}$.

7) Compute $\hat{q}_k$, $k = 1, 2, \cdots, \hat{K}$, by determining the eigenvalues of $\hat{\bar{\Psi}}_q$.

8) Calculate the direction estimates $\hat{\theta}_k$ from

$$\hat{\theta}_k = -\sin^{-1}\left\{\arg(\hat{q}_k)\frac{\lambda}{2\pi\delta}\right\}, \quad k = 1, 2, \cdots, \hat{K}. \tag{36}$$

IV. Simulation Results

We show below several examples that compare the performance of this estimator with that of the full polarization method in [1]. We also compare this estimator with ESPRIT estimators using only the x- or y-axis dipoles (i.e., a conventional ESPRIT array).

The results below are for an array with L = five pairs of crossed dipoles with a spacing δ between adjacent dipole pairs of a half-wavelength. All incident signals are assumed to have the same unit amplitude E_k. The signal-to-noise ratio (SNR) used in the simulations ($-10 \log_{10} \sigma^2$ dB) is 20 dB. The number of data samples taken at each dipole output is $N = 31$, and 50 Monte Carlo simulations were done for each case.

We begin with the case of a single linearly polarized signal ($\alpha = 0°$). Fig. 3 shows the variance (in decibels with respect to degrees squared) of the direction estimate $\hat{\theta}$ as a function of β when $\theta = 20°$. Four curves are shown, one for the full polarization method of [1], one for the angle-only method discussed above, and two for conventional ESPRIT estimators using only the x-dipoles or y-dipoles. Note that the performance of the angle-only method is almost the same as that of the full polarization method and is not sensitive to β. The performance of the conventional ESPRIT estimators using only the x- or y-axis dipoles, however, is sensitive to β. For example, when only the x-axis dipoles are used, the angle estimates deteriorate rapidly as β approaches 90° (as the signal becomes vertically polarized). Of course, the reason is that the signals on the x-axis dipoles approach zero as β approaches 90°. It is thus an advantage to use an array with elements responding to more than one polarization.

Fig. 4 shows another example of a single signal with $\beta = 0°$ and $\theta = 20°$. Fig. 4 shows the variance of the direction estimates as a function of the ellipticity angle α. Note again that performance of the angle-only method is almost the same as that of the full polarization method and is not sensitive to α. The performance of a conventional ESPRIT estimator using the x-axis dipoles is also not very sensitive to α, because the signals on the x-axis dipoles are never close to zero for any α. For an ESPRIT estimator using the y-axis dipoles, however, the variance blows up when the polarization approaches linear (near $\alpha = 0°$).

Next we consider a case where two signals arrive from $\theta_1 = 20°$ and $\theta_2 = 20° + \Delta\theta$, so $\Delta\theta$ is the angular separation between the two signals. We assume the corresponding ellipticity angles are $\alpha_1 = 45°$ and $\alpha_2 = 45° - \Delta\alpha$ and the orientation angles are $\beta_1 = \beta_2 = 0°$, so $\Delta\alpha$ is the only difference in polarization between the two signals. Fig. 5 shows the variance of $\hat{\theta}_1$ as a function of $\Delta\alpha$ for two values of $\Delta\theta$. The variance is obtained by taking the smaller of the two angle estimates as $\hat{\theta}_1$.

Note that the performance of the angle-only approach is not sensitive to $\Delta\alpha$, but the performance of the full polarization approach is. For small $\Delta\theta$ and large $\Delta\alpha$, the full polarization method yields better performance than the angle-only method. For large $\Delta\theta$ or small $\Delta\alpha$, the two methods yield similar performance. The poorer performance of the angle-only method when $\Delta\theta$ is small and $\Delta\alpha$ is large occurs because $\bar{\mathbf{A}}_L$ in (20) is ill-conditioned but $\mathbf{A}$ in (15) is not. The ill-conditioned $\bar{\mathbf{A}}_L$ makes it hard to resolve the columns of $\bar{\mathbf{A}}_L$ and thus to resolve the two closely spaced directions.

Note also from Fig. 5 that for a conventional ESPRIT estimator using the y-axis dipoles, the variance of $\hat{\theta}_1$ increases as $\Delta\alpha$ approaches 45° (linear polarization) even though $\Delta\alpha$ affects only the polarization of the signal from θ_2. The reason is that as $\Delta\alpha$ approaches 45°, the signals on the y-axis dipoles due to signal 2 go to zero, so the estimates of θ_2 become very poor. Because the smaller of the two estimated angles is taken to be $\hat{\theta}_1$, a large error in the angle estimate for signal 2 can cause the two estimated angles to be assigned incorrectly to $\hat{\theta}_1$ and $\hat{\theta}_2$. The result is a large increase in the computed variance of $\hat{\theta}_1$.

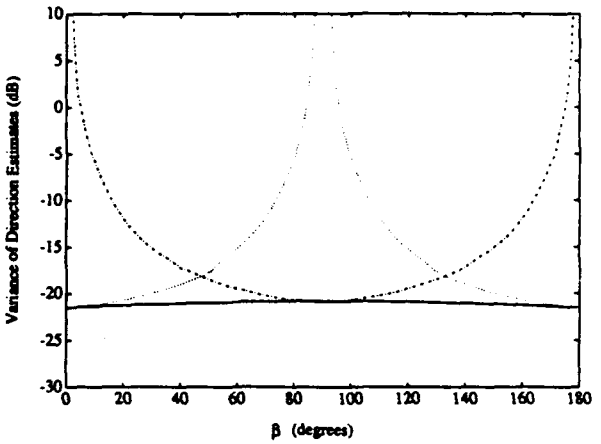


Fig. 3. Variance of $\hat{\theta}$ versus β for a linearly polarized signal ($\alpha = 0°$), $\theta = 20°$. Solid curve: full polarization method and angle-only method (the two curves are virtually the same for this case); dotted curve: x-axis dipoles only; dashdot curve: y-axis dipoles only.

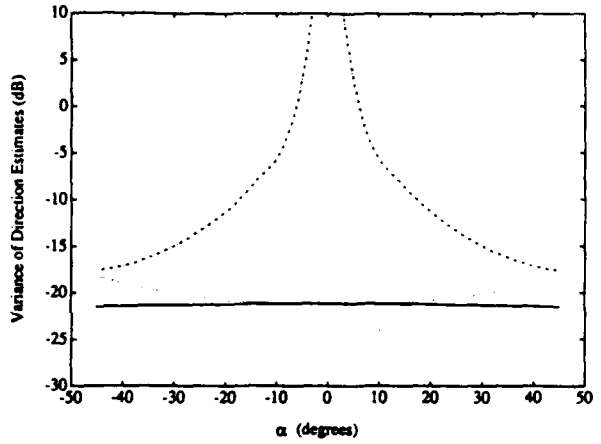


Fig. 4. Variance of $\hat{\theta}$ versus ellipticity α for $\beta = 0°$, $\theta = 20°$. Solid curve: full polarization method and angle-only method (the two curves are virtually the same for this case); dotted curve: x-axis dipoles only; dashdot curve: y-axis dipoles only.

Finally, we remark that the above results were obtained by assuming that the number of incident signals K is known. We found, however, that the MDL criterion provided accurate estimates of K for both the angle-only method and the full polarization method in all simulations. For a conventional ESPRIT estimator using only the x- or y-axis dipoles, however, estimates of K can be wrong for certain signal polarizations.

V. Conclusion

In a previous paper [1], we described how the ESPRIT algorithm can be used to estimate both the arrival directions and the polarizations of incoming plane waves with a uniform linear array of crossed dipoles. In this communication, we have described a simpler alternative approach that can be used to estimate only the arrival directions. This new approach requires approximately one eighth as many computations as the earlier method (the full polarization method). Simulation results show that the full polarization method yields better direction estimates than this new approach if

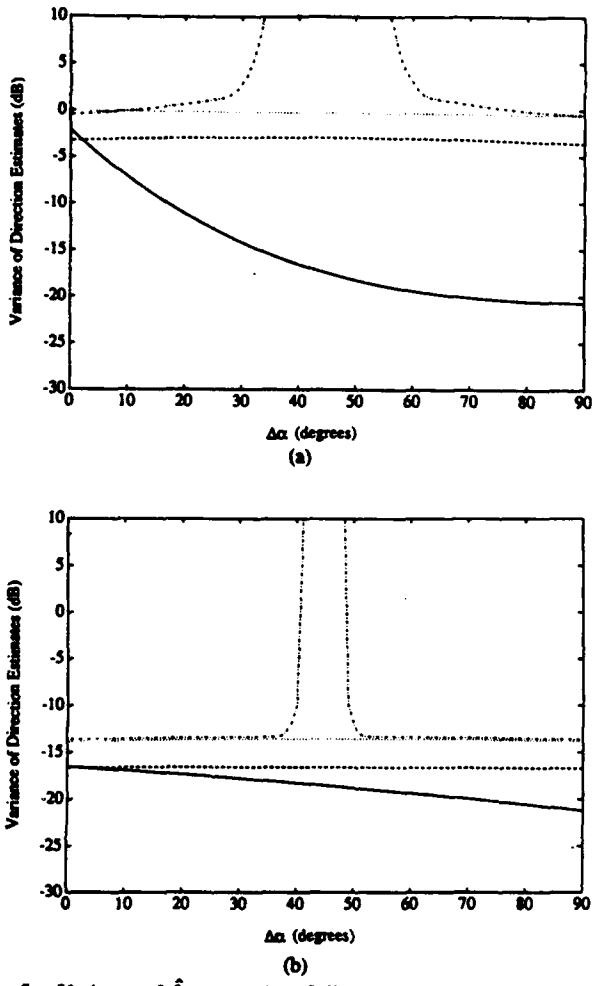


Fig. 5. Variance of $\hat{\theta}$ versus $\Delta\alpha$. Solid curve: full polarization method; dashed curve: angle-only method; dotted curve: x-axis dipoles only; dashdot curve: y-axis dipoles only. (a) $\Delta\theta = 4°$. (b) $\Delta\theta = 20°$.

the difference in the signal directions is small, but the difference in the polarizations is large. Otherwise, the two approaches have similar performance.

References

[1] J. Li and R.T. Compton, Jr., "Angle and polarization estimation using ESPRIT with a polarization sensitive array," *IEEE Trans. Antennas Propagat.*, vol. 39, pp. 1376–1383, Sept. 1991.

[2] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.

[3] C. A. Balanis, *Antenna Theory—Analysis and Design*. New York, Harper & Row, 1982.

[4] G. A. Deschamps, "Geometrical representation of the polarization of a plane electromagnetic wave," *Proc. IRE*, vol. 39, pp. 540–544, May 1951.

[5] S. U. Pillai, *Array Signal Processing*. New York: Springer-Verlag, 1989.

[6] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins University, 1989.

[7] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[8] G. H. Golub and C. F. V. Loan, "An analysis of the total least squares problem," *SIAM J. Numerical Analysis*, vol. 17, pp. 883–893, Dec. 1980.

A Microstrip Line on a Chiral Substrate

Michael S. Kluskens and Edward H. Newman

Abstract—**Right and left circular vector potentials are developed and used in a spectral-domain solution for a microstrip transmission line on a chiral substrate. These vector potentials have properties similar to those of the usual magnetic and electric vector potentials, except that they result in circular rather than linearly polarized fields, thereby simplifying field expansions in chiral media. The chiral microstrip line does not have bifurcated modes like other chiral guided wave structures; however, the chiral substrate causes a significant asymmetry in both the fields and currents.**

I. Introduction

This paper presents a spectral-domain Galerkin moment method (MM) solution for a microstrip transmission line on a chiral substrate. A chiral medium is a form of artificial dielectric consisting of chiral objects randomly embedded in a dielectric or other medium [1]. At optical frequencies, the chiral objects are molecules and the medium is called an isotropic optically active medium. At microwave frequencies, early research used conducting helices as a scale model for optical activity [2]. From this and later work, the constitutive relationships for chiral media have been shown to be the same as those for isotropic optically active media; therefore, the same notation is used [3, sec. 8.3].

A chiral medium is distinguished from other media in that right and left circularly polarized waves propagate through it with different phase velocities, even though it is a reciprocal and isotropic medium. For most chiral guided wave structures this property results in bifurcated modes [4]–[6], i.e., pairs of modes with the same cutoff frequency. The chiral microstrip line does not have bifurcated modes, and thus the dispersion curves are single valued. The primary effect of the chiral substrate is to generate asymmetric longitudinal and symmetric transverse fields. This effect could significantly alter the properties of microwave devices constructed on a chiral substrate.

Manuscript received March 1, 1991; revised July 8, 1991. This work was sponsored by the Joint Service Electronics Program under Contract N00014-78-C-0049 with the Ohio State University Research Foundation.

The authors are with the Department of Electrical Engineering, The Ohio State University, ElectroScience Laboratory, 1320 Kinnear Road, Columbus, OH 43212.

IEEE Log Number 9102813.

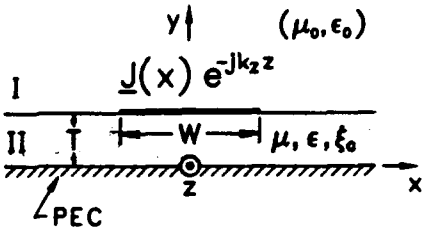


Fig. 1. Microstrip line on a grounded chiral substrate.

II. Theory

The constitutive relationships for a chiral medium can be written as

$$D = \epsilon_c E - j\mu\xi_c H \tag{1}$$

$$B = \mu H + j\mu\xi_c E \tag{2}$$

where $\epsilon_c = \epsilon + \mu\xi_c^2$, μ is the permeability, ϵ is the permittivity, and the pseudoscalar ξ_c is the chirality admittance of the medium ($e^{j\omega t}$).

Following the techniques used in [7], [8], the right (R) and left (L) circular vector potentials are defined as

$$R = \hat{a}\psi(k_R) \tag{3}$$

$$L = \hat{a}\psi(k_L) \tag{4}$$

where $\hat{a}$ is an arbitrary unit vector and $\psi(k)$ is a solution of the scalar wave equation $\nabla^2\psi(k) + k^2\psi(k) = 0$. The right and left circularly polarized electric fields are formed using

$$E_R = \nabla \times \left(R + \frac{1}{k_R} \nabla \times R \right) \tag{5}$$

$$E_L = \nabla \times \left(L - \frac{1}{k_L} \nabla \times L \right) \tag{6}$$

where the wave numbers k_R and k_L are given by

$$\left.\begin{matrix} k_R \\ k_L \end{matrix}\right\} = \omega\sqrt{\mu\epsilon_c} \pm \omega\mu\xi_c. \tag{7}$$

The corresponding magnetic fields are given by

$$\begin{pmatrix} H_R \\ H_L \end{pmatrix} = \frac{j}{\eta_c} \begin{pmatrix} E_R \\ -E_L \end{pmatrix} \tag{8}$$

where $\eta_c = \sqrt{\mu/\epsilon_c}$ is the chiral wave impedance. The right (or left) circular vector potential component R_y (or L_y) produces a right (or left) circular to y field RC_Y (or LC_Y), just as the magnetic vector potential component A_y produces a transverse magnetic to y field TM_Y.

The microstrip line is shown in Fig. 1, where the substrate has parameters (μ, ϵ, ξ_c) and thickness T. The microstrip line is W wide, infinitely thin, and perfectly conducting with a current distribution of $J(x)e^{-jk_z z}$. The region $y > T$ is free space, with parameters (μ_0, ϵ_0) and wave number $k_0 = \omega\sqrt{\mu_0\epsilon_0}$. In this region the fields may be expanded as the sum of TM_Y field and a TE_Y field using

$$\begin{pmatrix} A \\ F \end{pmatrix} = \frac{\hat{y}}{2\pi} \int_{-\infty}^{\infty} \begin{pmatrix} \tilde{A} \\ \tilde{F} \end{pmatrix} e^{-j(k_x x + k_y y + k_z z)}\, dk_x \tag{9}$$

where $k_y^2 = k_x^2 + k_z^2 - k_0^2$.

In the substrate, the fields are expanded in terms of right and left circular vector potentials. Individually, right or left circularly polarized fields can not satisfy the boundary condition of zero

tangential electric field on the ground plane at $y = T$. However, this boundary condition can be satisfied by a quasi-TM_Y field formed as the sum of a RC_Y field and a LC_Y field generated by the circular vector potentials [8]:

$$\begin{pmatrix} R_{Y,M} \\ L_{Y,M} \end{pmatrix} = \frac{\hat{y}}{2\pi} \int_{-\infty}^{\infty} \tilde{Q}_M \begin{pmatrix} \cos k_{y,R} y \\ -\cos k_{y,L} y \end{pmatrix} e^{-j(k_x x + k_z z)} dk_x \quad (10)$$

where $k_{y,R}^2 = k_x^2 + k_z^2 - k_R^2$ and $k_{y,L}^2 = k_x^2 + k_z^2 - k_L^2$. The resulting field is TM_Y if $\xi_c = 0$; hence the name quasi-TM_Y. Similarly, a quasi-TE_Y field can be formed using

$$\begin{pmatrix} R_{Y,E} \\ L_{Y,E} \end{pmatrix} = \frac{\hat{y}}{2\pi} \int_{-\infty}^{\infty} \tilde{Q}_E \begin{pmatrix} \frac{k_R}{k_{y,R}} \sin k_{y,R} y \\ \frac{k_L}{k_{y,L}} \sin k_{y,L} y \end{pmatrix} e^{-j(k_x x + k_z z)} dk_x. \quad (11)$$

The four unknown spectral functions $\tilde{A}$, $\tilde{F}$, $\tilde{Q}_M$, and $\tilde{Q}_E$ are determined by enforcing the boundary conditions at $y = T$ [8]. The fields E_x and E_z at the interface $y = T$ are presented below in terms of the even and odd components of the Fourier transforms of these fields generated by $\hat{x}$ and $\hat{z}$ polarized traveling wave line sources at $x = 0$, $y = T$:

$$\tilde{E}_{x,e}^{J_x} = \frac{1}{\Delta}\left[\left(k_x^2 \frac{k_y}{\omega\mu_0} + k_z^2 \frac{\omega\epsilon_0}{k_y}\right)(1-S) - \frac{j}{\eta_c}\left(k_x^2 U + k_z^2 V\right)\right] \quad (12)$$

$$\tilde{E}_{z,e}^{J_z} = \frac{1}{\Delta}\left[\left(k_x^2 \frac{\omega\epsilon_0}{k_y} + k_z^2 \frac{k_y}{\omega\mu_0}\right)(1-S) - \frac{j}{\eta_c}\left(k_x^2 V + k_z^2 U\right)\right] \quad (13)$$

$$\tilde{E}_{z,o}^{J_x} = \tilde{E}_{x,o}^{J_z} = \frac{k_x k_z}{\Delta}\left[\left(\frac{k_y}{\omega\mu_0} - \frac{\omega\epsilon_0}{k_y}\right)(1-S) + \frac{j}{\eta_c}(V-U)\right] \quad (14)$$

$$\tilde{E}_{x,o}^{J_x} = -\tilde{E}_{z,o}^{J_z} = j\frac{2k_x k_z}{\eta_c \Delta} G^- \quad (15)$$

$$\tilde{E}_{z,e}^{J_x} = \tilde{E}_{x,e}^{J_z} = j\frac{k_z^2 - k_x^2}{\eta_c \Delta} G^- \quad (16)$$

where

$$\Delta = \left(k_x^2 + k_z^2\right)\left\{\left[\frac{\epsilon_c}{\mu}(1+S) - \frac{\epsilon_0}{\mu_0}(1-S)\right] + j\left[\frac{k_2}{k_y}\frac{\epsilon_0}{\mu}U + \frac{k_y}{k_2}\frac{\epsilon_c}{\mu_0}V\right]\right\} \quad (17)$$

$$G^{\pm} = \frac{1}{2}\left(\frac{k_R}{k_{y,R}}\frac{k_{y,L}}{k_L} \pm \frac{k_{y,R}}{k_R}\frac{k_L}{k_{y,L}}\right)\sin k_{y,R}T \sin k_{y,L}T \quad (18)$$

$$S = \cos k_{y,R}T \cos k_{y,L}T - G^+ \quad (19)$$

$$U = \frac{k_{y,R}}{k_R}\sin k_{y,R}T \cos k_{y,L}T + \frac{k_{y,L}}{k_L}\sin k_{y,L}T \cos k_{y,R}T \quad (20)$$

$$V = \frac{k_R}{k_{y,R}}\sin k_{y,R}T \cos k_{y,L}T + \frac{k_L}{k_{y,L}}\sin k_{y,L}T \cos k_{y,R}T \quad (21)$$

with $k_2 = \omega\sqrt{\mu\epsilon_c} = (k_R + k_L)/2$. For example, the E_x field generated by the surface current $J_z(x)$ is given by

$$E_x^{J_z}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\tilde{E}_{x,e}^{J_z}(k_x) + \tilde{E}_{x,o}^{J_z}(k_x)\right] \cdot \tilde{J}_z(k_x) e^{-j(k_x x + k_z z)} dk_x. \quad (22)$$

In a conventional achiral microstrip line $\tilde{E}_{x,o}^{J_x}$, $\tilde{E}_{z,o}^{J_z}$, $\tilde{E}_{z,e}^{J_x}$, and $\tilde{E}_{x,e}^{J_z}$ are zero, causing $J_x(x)$ and $J_z(x)$ to be even and odd functions of x, respectively. However, this is not true for a chiral microstrip line, thereby requiring a set of even and odd basis functions.

III. Moment Method Solution

The J_x and J_z currents for the MM solution are expanded as

$$J_x(x) = \sum_{n=0}^{N_x} I_{x,n} J_{x,n}(x) \quad (23)$$

$$J_z(x) = \sum_{n=0}^{N_z} I_{z,n} J_{z,n}(x) \quad (24)$$

where $I_{x,n}$ and $I_{z,n}$ are the unknown coefficients. The basis functions $J_{x,n}$ and $J_{z,n}$ are Chebyshev polynomials weighted by the edge conditions [9]–[13]:

$$J_{x,n}(x) = \frac{4}{\pi W}\frac{U_n(2x/W)}{n+1}\sqrt{1-(2x/W)^2} \quad (25)$$

$$J_{z,n}(x) = \frac{2}{\pi W} T_n(2x/W) \Big/ \sqrt{1-(2x/W)^2} \quad (26)$$

where $T_n(x)$ and $U_n(x)$ are Chebyshev polynomials of the first and second kinds, respectively. The Fourier transforms of these basis functions are [14, sec. 6.671]:

$$\tilde{J}_{x,n}(k_x) = 2j^n \frac{J_{n+1}(k_x W/2)}{k_x W/2} \quad (27)$$

$$\tilde{J}_{z,n}(k_x) = j^n J_n(k_x W/2) \quad (28)$$

where $J_n(x)$ is an nth-order Bessel function. The MM can then be applied to enforce the boundary condition of zero tangential electric field on the microstrip line. In block matrix form, the MM equation is

$$\begin{bmatrix} Z_{xx} & Z_{xz} \\ Z_{zx} & Z_{zz} \end{bmatrix}\begin{bmatrix} I_x \\ I_z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (29)$$

where $I_x = [I_{x,0} \cdots I_{x,N_x}]^T$ and $I_z = [I_{z,0} \cdots I_{z,N_z}]^T$.

In the xz block

$$Z_{xz} = \begin{bmatrix} Z_{xz(0,0)} & \cdots & Z_{xz(0,N_z)} \\ \vdots & \ddots & \vdots \\ Z_{xz(N_x,0)} & \cdots & Z_{xz(N_x,N_z)} \end{bmatrix} \quad (30)$$

$$Z_{xz(m,n)} = -\int_{-\infty}^{\infty} \tilde{E}_x^{J_z}(k_x) \tilde{J}_{z,n}(k_x) \tilde{J}_{x,m}(-k_x) dk_x. \quad (31)$$

Impedance elements in the remaining blocks are given similarly.

The propagation constants of the modes are found as the roots of the determinant of the impedance matrix given in (29). For a given propagation constant the fields in any region and the current distribution may be found using the equations presented in the previous section.

IV. Numerical Results

This section presents numerical results demonstrating the accuracy of the MM solution, and some effects of chirality on a microstrip transmission line. All currents are normalized to

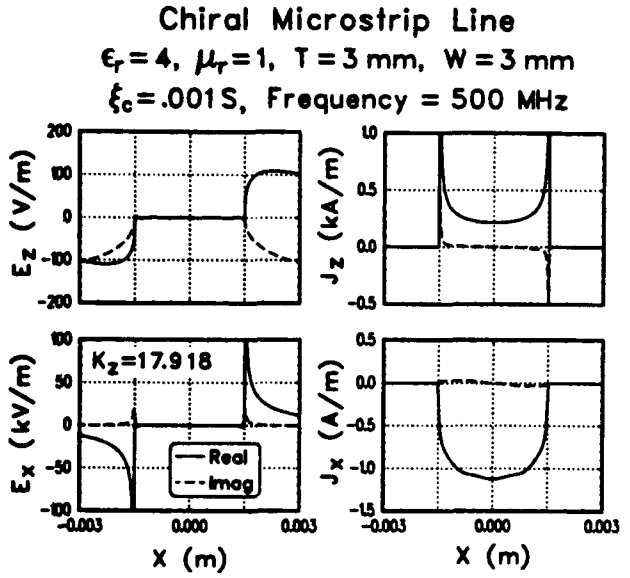


Fig. 2. Fields and currents at $y = T$ for a chiral microstrip solved using ten J_z modes and ten J_x modes.

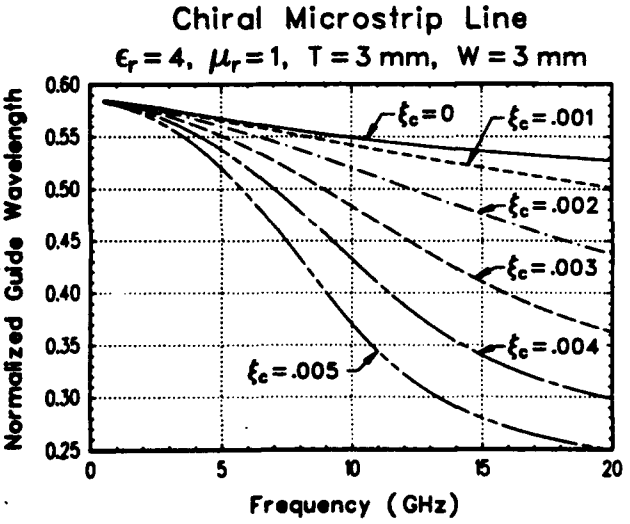


Fig. 3. Normalized guide wavelength (λ_g / λ_0) versus frequency for the fundamental mode of chiral and achiral microstrip lines, for a range of chiral parameters in Siemens.

$I_{z,0} = 1$, since the microstrip current can only be found to within a constant. In the figures the real part of the current and fields is shown as a solid line, and the imaginary part as a dashed line.

Fig. 2 shows the electric fields and currents at the interface $y = T$ for a MM solution using ten longitudinal and ten transverse basis functions. The left-hand graphs show that the fields satisfy the boundary condition of zero tangential electric field on the microstrip line. The corresponding currents are shown in the right-hand graphs. The even transverse current component, which occurs solely because of the chirality, is significantly larger than the odd transverse current component.

The dispersion curve shown in Fig. 3 shows the normalized guide wavelength (λ_g / λ_0) for the fundamental mode of a chiral microstrip line, for a range of chiral parameters. The case $\xi_c = 0$ corresponds to an achiral line. Fig. 3 shows that the propagation constant is not significantly affected unless the chiral parameter is a significant percentage of the maximum value set in [15] of $\xi_{c,\max} = \sqrt{\epsilon / \mu}$, which in this case is 0.0053 S.

References

[1] D. L. Jaggard, A. R. Mickelson, and C. H. Papas, "On electromagnetic waves in chiral media," *Appl. Phys.*, vol. 18, pp. 211–216, 1979.

[2] I. Tinoco, Jr. and M. P. Freeman, "The optical activity of oriented copper helices. I. Experimental," *J. Phys. Chem.*, vol. 61, pp. 1196–1200, Sept. 1957.

[3] E. J. Post, *Formal Structure of Electromagnetics*. Amsterdam: North-Holland, 1962.

[4] C. Eftimiu and L. W. Pearson, "Guided electromagnetic waves in chiral media," *Radio Science*, vol. 24, pp. 351–359, May/June 1989.

[5] P. Pelet and N. Engheta, "The theory of chirowaveguides," *IEEE Trans. Antennas Propagat.*, vol. AP-38, pp. 90–98, Jan. 1990.

[6] J. A. M. Svedin, "Propagation analysis of chirowaveguides using the finite-element method," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 1488–1496, Oct. 1990.

[7] M. S. Kluskens and E. H. Newman, "Scattering by a multilayer chiral cylinder," *IEEE Trans. Antennas Propagat.*, vol. AP-39, pp. 91–96, Jan. 1991.

[8] M. S. Kluskens, "Method of moments analysis of scattering by chiral media," Ph.D. dissertation, The Ohio State University, Columbus, June 1991.

[9] R. E. Collin, *Field Theory of Guided Waves*. New York: IEEE Press, 1991.

[10] T. Kitazawa and Y. Hayashi, "Propagation characteristics of striplines with multilayered anisotropic media," *IEEE Trans. Microwave Theory Tech.*, vol. 31, pp. 429–433, June 1983.

[11] F. Medina, M. Horno, and H. Baudrand, "Generalized spectral analysis of planar lines on layered media including uniaxial and biaxial dielectric substrates," *IEEE Trans. Microwave Theory Tech.*, vol. 37, pp. 504–511, Mar. 1989.

[12] M. Kobayashi and T. Iijima, "Frequency-dependent characteristics of current distributions on microstrip," *IEEE Trans. Microwave Theory Tech.*, vol. 37, pp. 799–801, Apr. 1989.

[13] Y. Yuan and D. P. Nyquist, "Full-wave perturbation theory based upon electric field integral equations for coupled microstrip transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 1576–1584, Nov. 1990.

[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Orlando, FL: Academic, 1980.

[15] N. Engheta and D. L. Jaggard, "Electromagnetic chirality and its applications," *IEEE Antennas and Propagation Society Newsletter*, vol. 30, pp. 6–12, Oct. 1988.

Journal of Electromagnetic Waves and Applications, Vol. 6, No. 5/6, 721-731, 1992

Scattering by a Chiral Cylinder of Arbitrary Cross Section in the Presence of a Half-Plane

M. S. Kluskens and E. H. Newman

Department of Electrical Engineering
The Ohio State University
ElectroScience Laboratory
1320 Kinnear Rd.
Columbus, Ohio 43212

Abstract– An integral equation and method of moments/Green's function solution to the problem of scattering by a chiral cylinder of arbitrary cross section in the presence of a perfectly conducting half-plane is presented. The volume equivalence theorem for chiral media is used to formulate a pair of coupled vector integral equations for the equivalent electric and magnetic volume polarization currents representing the chiral cylinder. The presence of the half-plane is accounted for by including the half-plane Green's function in the kernel of the integral equations, and efficient techniques for accurately evaluating the integrals in this Green's function are presented. Numerical results illustrate that a chiral cylinder surrounding the half-plane edge can significantly modify the scattering from the edge. The chiral cylinder is also seen to produce significant cross-polarized scattered fields, which are a direct result of the rotation of field polarization in a chiral medium.

I. INTRODUCTION

This paper presents a pulse-basis point-matching method of moments (MM)/ Green's function solution [1] to the two-dimensional problem of scattering by an isotropic and inhomogeneous chiral cylinder of arbitrary cross section in the presence of a perfectly conducting half-plane. The solution allows the efficient analysis of the effects of a chiral coating on the scattering from an ideal knife edge. This technique can be used to study the use of chiral media to modify or control edge diffraction, which could have application to reflector antennas or to antennas on finite ground planes. Also, the MM solution can serve as a reference solution for asymptotic or approximate methods.

A chiral medium is a reciprocal medium characterized by different phase velocities for right and left circularly polarized waves. In a lossless chiral medium, a linearly polarized wave undergoes a rotation of its polarization as it propagates. These are the same properties as an isotropic optically active media; therefore, the same constitutive relationships can be used [2, sec. 8.3]. For chiral cylinders, these properties result in a coupling of the TM and TE polarizations. A partial list of references describing chiral media is given by [3–9].

Following the authors' previous work [10,11,12], the present solution uses the chiral volume equivalence theorem [10] to replace the chiral cylinder by free space and equivalent electric and magnetic volume polarization currents (J, M). These

polarization currents are formulated as the solution to a pair of coupled vector integral equations. The electric surface currents on the half-plane are not explicitly included as unknowns in the MM solution. Instead, their effects are exactly accounted for by including the half-plane Green's function in the kernel of the integral equation [13, sec. 8.3]. Efficient techniques for accurately evaluating the integrals in this Green's function are presented in the appendix. These coupled vector integral equations for $(\overline{J}, \overline{M})$ are equivalent to six coupled scalar integral equations for (J_x, J_y, J_z) and (M_x, M_y, M_z), which are solved using a pulse-basis point-matching MM solution. It is noted that this problem could be formulated in terms of surface currents on the chiral cylinder [5,14]. The main advantage of the volume formulation is that it can more easily treat inhomogeneous media.

Numerical results, including echo width and internal fields, are presented for the scattering by two geometries, a chiral slab and a double-wedge at the tip of the perfectly conducting half-plane. These results are compared with the fields produced by similar achiral bodies in the presence of the half-plane and the bare half-plane. In addition, the cross-polarized fields, produced by the coupling between the TM and TE polarizations in a chiral medium, are shown to be of comparable magnitude to the co-polarized fields.

II. DERIVATION OF THE INTEGRAL EQUATIONS

This section develops a set of coupled integral equations for the equivalent electric and magnetic currents representing a chiral cylinder in the presence of a perfectly conducting half-plane. In this paper, all fields and currents are considered to be time harmonic with the $e^{j\omega t}$ time dependence suppressed. The constitutive relationships for a chiral medium [2, sec 8.3] can be written as

$$\overline{D} = \epsilon \overline{E} - j\xi_c \overline{B} \tag{1}$$

$$\overline{H} = \frac{1}{\mu} \overline{B} - j\xi_c \overline{E} \tag{2}$$

where μ is the permeability, ϵ is the permittivity, and the pseudoscalar ξ_c [2, p. 176] is the chirality admittance of the medium. If μ, ϵ, or ξ_c are complex the media is lossy. In an inhomogeneous chiral medium μ, ϵ, and ξ_c are functions of position. If $\xi_c = 0$, then (1) and (2) reduce to the constitutive relations for an achiral medium. To simplify the following developments, (1) and (2) can be written as

$$\overline{D} = \epsilon_c \overline{E} - j\mu\xi_c \overline{H} \tag{3}$$

$$\overline{B} = \mu \overline{H} + j\mu\xi_c \overline{E} \tag{4}$$

where the effective permittivity, ϵ_c, of the chiral medium is defined by

$$\epsilon_c = \epsilon + \mu\xi_c^2 \tag{5}$$

In the original problem of Fig. 1a, the impressed currents $\left(\overline{J}^i, \overline{M}^i\right)$ radiate the total fields $\left(\overline{E}^T, \overline{H}^T\right)$ in a medium which is free space except for the perfectly conducting half-plane and a chiral cylinder with constitutive parameters (μ, ϵ, ξ_c)

confined to the region R. The chiral cylinder may be lossy and/or inhomogeneous. As shown in Fig 1b., the chiral volume equivalence theorem [10] is used to replace the chiral cylinder by free space and the electric and magnetic volume polarization currents

$$\overline{J} = j\omega(\epsilon_c - \epsilon_0)\overline{E}^T + \omega\mu\xi_c\overline{H}^T \tag{6}$$

$$\overline{M} = j\omega(\mu - \mu_0)\overline{H}^T - \omega\mu\xi_c\overline{E}^T \tag{7}$$

confined to the region R, where (μ_0, ϵ_0) are the constitutive parameters of free space.

In the equivalent problem of Fig. 1b, the total fields $\left(\overline{E}^T, \overline{H}^T\right)$ at any point in space are given by

$$\overline{E}^T = \overline{E}^i + \overline{E}^J + \overline{E}^M \tag{8}$$

$$\overline{H}^T = \overline{H}^i + \overline{H}^J + \overline{H}^M \tag{9}$$

where $\left(\overline{E}^i, \overline{H}^i\right)$, $\left(\overline{E}^J, \overline{H}^J\right)$, and $\left(\overline{E}^M, \overline{H}^M\right)$ are the fields radiated by $\left(\overline{J}^i, \overline{M}^i\right)$, $\overline{J}$, and $\overline{M}$, respectively, in the presence of the half-plane. Then, substituting (8) and (9) into (6) and (7) yields

$$-\overline{E}^J - \overline{E}^M - c_E\left[\overline{H}^J + \overline{H}^M\right] + a_E\overline{J} = \overline{E}^i + c_E\overline{H}^i \quad \text{in } R \tag{10}$$

$$-\overline{H}^J - \overline{H}^M - c_M\left[\overline{E}^J + \overline{E}^M\right] + a_M\overline{M} = \overline{H}^i + c_M\overline{E}^i \quad \text{in } R \tag{11}$$

where c_E, c_M, a_E, and a_M are defined by

$$c_E = \frac{\omega\mu\xi_c}{j\omega(\epsilon_c - \epsilon_0)} \tag{12}$$

$$c_M = -\frac{\omega\mu\xi_c}{j\omega(\mu - \mu_0)} \tag{13}$$

$$a_E = \frac{1}{j\omega(\epsilon_c - \epsilon_0)} \tag{14}$$

$$a_M = \frac{1}{j\omega(\mu - \mu_0)} \tag{15}$$

Equations (10) and (11) can be considered to be coupled integral equations for $(\overline{J}, \overline{M})$ since the fields $\left(\overline{E}^J, \overline{H}^J\right)$ and $\left(\overline{E}^M, \overline{H}^M\right)$ can be written as integrals, over the region R containing the chiral cylinder, of $\overline{J}$ and $\overline{M}$, respectively, dotted into the appropriate half-plane dyadic Green's function [13, sec 8.3]. For example, $\overline{E}^M$ could be written as

$$\overline{E}^M(\overline{r}) = \int_R \overline{M}(\overline{r}') \cdot \overline{\overline{G}}(\overline{r}, \overline{r}')\, ds' \tag{16}$$

where $\overline{r}'$ is the source point, $\overline{r}$ is the field point, and $\overline{\overline{G}}$ is the electric field dyadic Green's function for a magnetic line source radiating in the presence of a perfectly conducting half-plane.

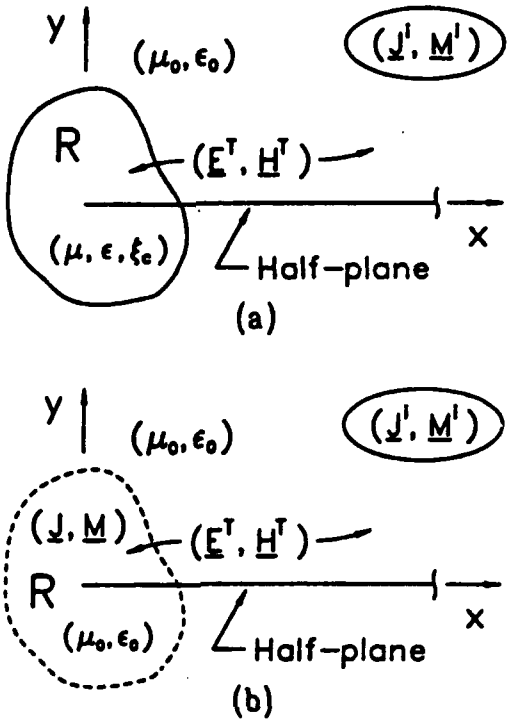


Figure 1. (a) Original problem: $\left(\overline{J}^i, \overline{M}^i\right)$, half-plane and chiral scatterer. (b) Equivalent problem: scatterer replaced by free space and $\left(\overline{J}, \overline{M}\right)$.

Equations (10) and (11) are then solved using a pulse-basis point-matching MM solution. The MM solution is notationally identical to that presented in [10] for a chiral cylinder in free space, except that all fields are evaluated in the presence of the half-plane instead of free space. Since the half-plane Green's function can be written as the sum of the free space Green's function plus a correction term which accounts for energy scattered by the half-plane [12,15], the elements in the MM matrix equation can be written as those for the chiral cylinder in free space plus a correction term which accounts for the presence of the half-plane. The free space components can be evaluated using the methods described in [10,11]. The evaluation of the correction term requires the numerical evaluation of a class of integrals. The efficient evaluation of these integrals is discussed in the appendix.

III. NUMERICAL RESULTS

This section presents numerical results for TM_z and TE_z scattering by two geometries, a chiral slab on a half-plane and a double-wedge covering the tip of a half-plane. All data is at a frequency of 300 MHz. The chiral parameters used in this section obey the limit set in [16] of $|\xi_c| < \sqrt{\epsilon/\mu}$. The literature contains a

limited amount of numerical data for scattering by chiral objects. Some numerical data is available for chiral spheres [17,18] and spheroids [19]. In a recent paper, the authors presented an eigenfunction solution for multilayer chiral cylinders [20].

Figure 2 shows the backscatter echo width pattern for a TM_z plane wave $\left(E_z^i = e^{-jk_0(x\cos\phi^i + y\sin\phi^i)}\right)$ incident upon a lossless chiral slab at the tip of a perfectly conducting half-plane. The slab is 1 meter wide and 0.2 meters thick, with parameters of $\epsilon_r = 4.0$, $\mu_r = 1.5$, and $\xi_c = 0.002\,\mathrm{A/V}$. For comparison, the echo widths of the achiral slab and of the bare half-plane are also shown. Although the half-plane is the dominant scatterer, the presence of the chiral slab does produce a significant change to the echo width. In particular, the chirality produces a cross-polarized component to the echo width which, in the region $\phi^i \leq 120°$, is of comparable magnitude to the co-polarized echo width. This cross-polarized field is a direct result of the rotation of polarization which occurs in a chiral medium. The magnitude of the internal fields along the center line ($y = 0.1\,\mathrm{m}$) of the slab are shown in Fig. 3 for $\phi^i = 60°$. Again, the cross-polarized fields (E_x, E_y, H_z) are a result of the rotation of field polarization in the chiral medium, and are of comparable magnitude to the co-polarized fields (E_z, H_x, H_y).

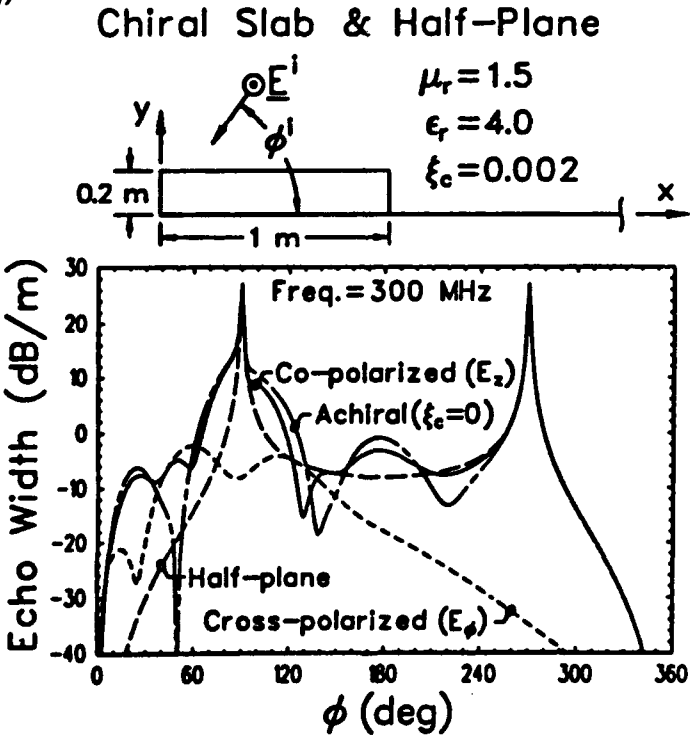


Figure 2. The co-polarized and cross-polarized backscatter ($\phi = \phi^i$) echo width of a lossless chiral slab at the tip of a perfectly conducting half-plane.

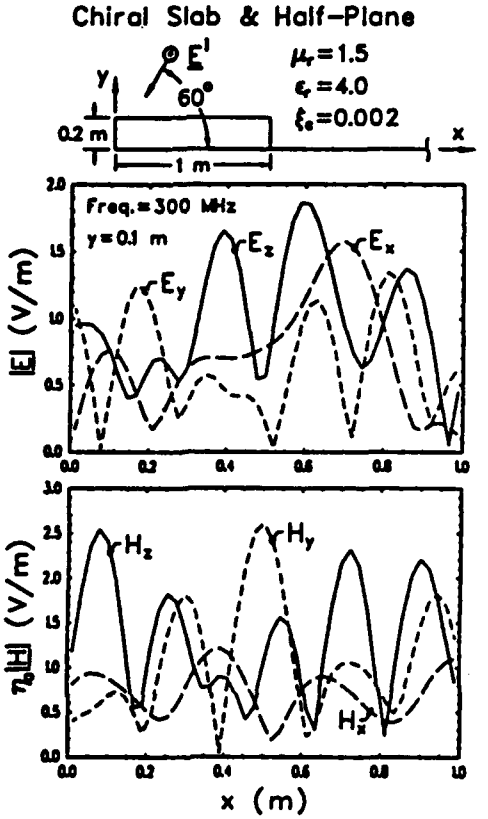


Figure 3. Internal fields along $y = 0.1$ meters for the chiral slab and half-plane geometry show in Fig. 2 with a TM $_Z$ incident wave from 60°.

Figure 4 shows the bistatic scattering from a perfectly conducting half-plane with a lossy chiral double-wedge covering the tip. The double-wedge is 2 meters wide and has a maximum thickness of 0.08 meters at the center. The upper graph shows the co-polarized and cross-polarized bistatic echo widths for a TM $_Z$ wave incident from 180° ($E_z^i = e^{-jk_0x}$). The bistatic echo width for an identical achiral ($\xi_c = 0$) double-wedge and for the bare half-plane is also shown for comparison. The lower graph displays the same data for a TE $_Z$ wave incident from 0° ($H_z^i = e^{jk_0x}$). In each case, the chiral double wedge causes a significant modification to the scattering from the half-plane. For example, for the TM $_Z$ case, the chiral wedge reduces the edge on backscatter ($\phi = 180°$) echo width from -8 dB/m to about -60 dB/m, while the achiral value is about -16 dB/m. The chirality also produces a significant cross-polarized component for both the TM $_Z$ and TE $_Z$ cases. Again, these cross-polarized fields are a direct result of the rotation of polarization in chiral medium.

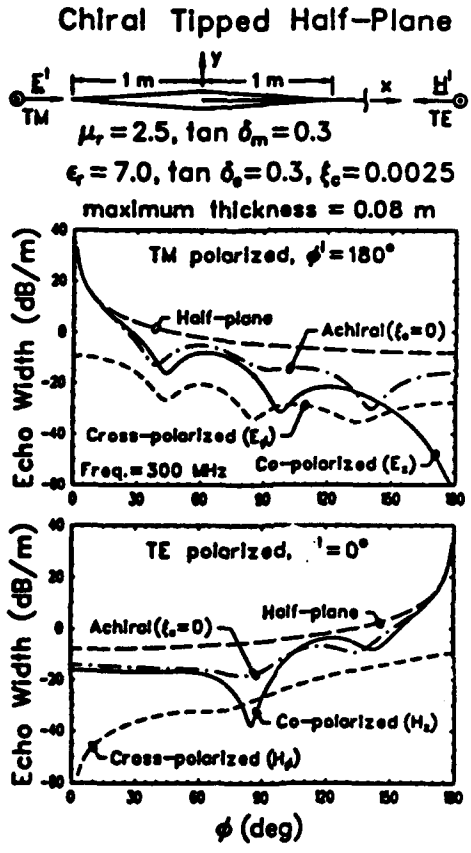


Figure 4. The co-polarized and cross-polarized bistatic echo width of a perfectly conducting half-plane with a chiral double-wedge at the tip.

IV. SUMMARY

This paper has presented an integral equation MM/Green's function solution to the problem of TM_z and TE_z scattering by an inhomogeneous chiral cylinder of arbitrary cross section in the presence of a perfectly conducting half-plane. The volume equivalence theorem for chiral media was used to formulate a pair of coupled vector integral equations for the equivalent electric and magnetic volume polarization currents representing the chiral cylinder. These equations were solved using the MM. The numerical data showed that the chirality of a simple slab at the tip of a perfectly conducting half-plane can significantly modify the scattering from the half-plane and also generate cross-polarized scattered and internal fields as large or larger than the co-polarized fields.

APPENDIX: EVALUATION OF THE INTEGRALS IN THE HALF-PLANE GREEN'S FUNCTION

The Green's function for a $\hat{z}$ polarized line source in the presence of a perfectly conducting half-plane involves the integrals [13, 8.3]

$$S_i(\alpha,\beta) = \int_a^\infty \frac{e^{-ju^2}}{(u^2+\beta^2)^{i-1/2}}\,du \tag{17}$$

for $i = 1, 2$. This appendix will present new and efficient techniques for accurately evaluating these integrals. For a lossless ambient medium α and β are real numbers. The range of α that needs to be explicitly considered can be restricted to $\alpha \geq 0$ using [12]

$$S_i(\alpha,\beta) = S_{i0}(\beta) - S_i(-\alpha,\beta) \tag{18}$$

where $i = 1, 2$ and the $S_{i0}(\beta)$ are defined by

$$S_{10}(\beta) = S_1(-\infty,\beta) = -j\frac{\pi}{2}e^{j\beta^2/2}H_0^{(2)}\left(\frac{\beta^2}{2}\right) \tag{19}$$

$$S_{20}(\beta) = S_2(-\infty,\beta) = -j\frac{\pi}{2}e^{j\beta^2/2}\left[H_0^{(2)}\left(\frac{\beta^2}{2}\right) + jH_1^{(2)}\left(\frac{\beta^2}{2}\right)\right] \tag{20}$$

Method I: Numerical Integration ($|\alpha| > 1.6$, all β)

This section presents a numerical integration technique for S_1 and S_2 requiring far fewer points then previous approaches [11,12,15]. Using the change of variables $u^2 = x$ and $x = \alpha^2 - jt$, where the second change of variables corresponds to changing the integration path for x to the contour $x = \alpha^2$ to $-j\infty + \alpha^2$ to ∞, (17) can be rewritten as

$$S_i(\alpha,\beta) = -j\frac{e^{-j\alpha^2}}{2}\int_0^\infty \frac{e^{-t}}{(\alpha^2 - jt + \beta^2)^{i-1/2}}\frac{dt}{\sqrt{\alpha^2 - jt}} \tag{21}$$

where the portion of the path at infinity does not contribute to the integral. These integrals are now in a form suitable for Gauss-Laguerre Quadrature [21, sec. 25.4.45], which is very efficient and requires only two integration points for $|\alpha| > 8$ to achieve six-digit accuracy.

Method II: Small Argument Approximations ($|\alpha| < 1.6$, $0 < \beta^2 < 7$)

This section presents efficient small argument approximations for the S_i integrals when α and β are small. The $\beta = 0$ case can be done analytically and is not discussed here. In this region the integrals are evaluated using

$$S_i(\alpha,\beta) = \frac{1}{2}S_{i0}(\beta) - \text{sign}(\alpha)\,\Delta S_i(0,|\alpha|,\beta) \tag{22}$$

where $i = 1, 2$, the S_{i0} are defined in equations (19) and (20), and the ΔS_i are

defined by

$$\Delta S_i(0,\alpha,\beta) = \int_0^{\alpha} \frac{e^{-ju^2}}{(u^2+\beta^2)^{i-1/2}} du \tag{23}$$

The ΔS_i can be evaluated using the Taylor series expansion for e^{-ju^2} and integration by parts, which produces

$$\Delta S_i(0,|\alpha|,\beta) = j^{1-i} a_{i,1} \ln\left(\left|\frac{\alpha}{\beta}\right| + \sqrt{\left(\frac{\alpha}{\beta}\right)^2 + 1}\right) + |\alpha|\left(\alpha^2+\beta^2\right)^{3/2-i}\left(j^{2i-3}\frac{b_{i,3-i}}{2^{2-i}} + \beta^{2(1-i)}\right) \tag{24}$$

where $a_{i,1}$ and $b_{i,n}$ are computed using reverse recursion and the relations

$$a_{i,n} = 1 + j\frac{n-\frac{3}{2}+i}{n(n-1+i)}\beta^2 a_{i,n+1} \tag{25}$$

$$b_{i,n} = a_{i,n} - j\frac{n-1}{n(n-1+i)}\alpha^2 b_{i,n+1}, \quad \text{for } n > 1 \tag{26}$$

$$b_{2,1} = a_{2,1} + j\frac{1}{4}\alpha^2 b_{2,2} \tag{27}$$

which are valid only for $i = 1, 2$.

Method III: Small/Large Argument Approximations ($|\alpha| < 1.6, \beta^2 \geq 7$)

This section presents efficient small/large argument approximations for the S_i integrals when $\alpha < \beta$ and β is large. Using (22) ΔS_i is evaluated using the binomial expansion for $(u^2+\beta^2)^{1/2-i}$ and integration by parts, which produces

$$\Delta S_i(0,\alpha,\beta) = \frac{1}{(\beta^2)^{i-1/2}}\left[c_{i,1}F(\alpha) - j\frac{i-\frac{1}{2}}{2}\frac{\alpha}{\beta^2}d_{i,2}\,e^{-j\alpha^2}\right] \tag{29}$$

where $c_{i,1}$ and $d_{i,2}$ are computed using reverse recursion and the relations

$$c_{i,n} = 1 + j\frac{\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}+i\right)}{n}\frac{1}{\beta^2}c_{i,n+1} \tag{30}$$

$$d_{i,n} = c_{i,n} - \frac{n-\frac{3}{2}+i}{n}\frac{\alpha^2}{\beta^2}d_{i,n+1} \tag{31}$$

$F(\alpha)$ is the Fresnel integral given by

$$F(\alpha) = \int_0^{\alpha} e^{-ju^2} du \tag{32}$$

ACKNOWLEDGEMENTS

This work was sponsored by the Joint Service Electronics Program under Contract N00014-78-C-0049 with the Ohio State Univ. Research Foundation, and by a grant from The Ohio Supercomputer Center.

The Guest Editor thanks X. Sun and one anonymous reviewer for reviewing the paper.

REFERENCES

1. Newman, E. H., "An overview of the hybrid MM/Green's function method in electromagnetics," *Proc. IEEE*, Vol. 76, 270–282, March 1988.
2. Post, E. J., *Formal Structure of Electromagnetics*, North-Holland, Amsterdam, 1962.
3. Weiglhofer, W. S., "Isotropic chiral media and scalar Hertz potentials," *J. Phys. A*, Vol. 21, 2249–2251, 1988.
4. Jaggard, D. L., X. Sun, and N. Engheta, "Canonical sources and duality in chiral media," *IEEE Trans. Antennas Propagat.*, Vol. AP-36, 1007–1013, July 1988.
5. Lakhtakia, A., V. V. Varadan, and V. K. Varadan, "Field equations, Huygens's principle, integral equations, and theorems for radiation and scattering of electromagnetic waves in isotropic chiral media," *J. Opt. Soc. Amer. A*, Vol. 5, 175–184, Feb. 1988.
6. Varadan, V. V., A. Lakhtakia, and V. K. Varadan, "On the equivalence of sources and duality of fields in isotropic chiral media," *J. Phys. A*, Vol. 20, 6259–6264, 1987.
7. Varadan, V. K., A. Lakhtakia, and V. V. Varadan, "Radiated potentials and fields in isotropic chiral media," *J. Phys. A*, Vol. 20, 4697–4702, 1987.
8. Bassiri, S., N. Engheta, and C. H. Papas, "Dyadic Green's function and dipole radiation in chiral media," *Alta Freq.*, Vol. LV, Mar.–Apr. 1986.
9. Jaggard, D. L., A. R. Mickelson, and C. H. Papas, "On electromagnetic waves in chiral media," *Appl. Phys.*, Vol. 18, 211–216, 1979.
10. Kluskens, M. S., and E. H. Newman, "Scattering by a chiral cylinder of arbitrary cross section," *IEEE Trans. Antennas Propagat.*, Vol. AP-38, 1448–1455, Sept. 1990.
11. Newman, E. H., "TM and TE scattering by a dielectric/ferrite cylinder in the presence of a half-plane," *IEEE Trans. Antennas Propagat.*, Vol. AP-34, 804–813, June 1986.
12. Newman, E. H., "TM scattering by a dielectric cylinder in the presence of a half-plane," *IEEE Trans. Antennas Propagat.*, Vol. AP-33, 773–782, July 1985.
13. Bowman, J. J., T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes*, Hemisphere, New York, 1987.
14. Rojas, R., "Integral equations for the scattering by three dimensional inhomogeneous chiral bodies," *J. Electro. Waves Applic.*, Vol. 6, No. 5/6, 787–804, 1992.
15. Newman, E. H., "The fields of a traveling wave line source in the vicinity of a half-plane," *IEEE Trans. Antennas Propagat.*, Vol. AP-35, 866–870, July 1987.
16. Engheta, N. and D. L. Jaggard, "Electromagnetic chirality and Its applications," *IEEE Antennas and Propagation Society Newsletter*, Vol. 30, 6–12, Oct. 1988.
17. Bohren, C. F., "Light scattering by an optically active sphere," *Chem. Phys. Lett.*, Vol. 29, 458–462, Dec. 1974.
18. Bohren, C. F., "Scattering of electromagnetic waves by an optically active spherical shell," *J. Chem. Phys.*, Vol. 62, 1566–1571, Feb. 1975.
19. Lakhtakia, A., V. K. Varadan, and V. V. Varadan, "Scattering and absorption characteristics of lossy dielectric, chiral, nonspherical objects," *Applied Optics*, Vol. 24, 4146–4154, Dec. 1985.
20. Kluskens, M. S. and E. H. Newman, "Scattering by a multilayer chiral cylinder," *IEEE Trans. Antennas Propagat.*, Vol. AP-39, 91–96, Jan. 1991.

21. Abramowitz, M. and I. A. Stegun, *Handbook of Mathematical Functions*, NBS, Washington, D.C., 1964.

Michael S. Kluskens was born in Chicago, Illinois on June 26, 1962. He received his BS and MS degrees in electrical engineering from Michigan Technological University in 1984 and 1985, respectively. Since 1986, he has been employed as a Graduate Research Associate at the ElectroScience Laboratory, Department of Electrical Engineering, the Ohio State University, and is currently working towards the Ph.D. degree. His research interest is in antennas and computational electromagnetics with emphasis on the Method of Moments. He is a member of Tau Beta Pi and Phi Kappa Phi.

Edward H. Newman was born in Cleveland, Ohio on July 9, 1946. He received the BSEE, MS, and Ph.D. degrees in electrical engineering from the Ohio State University in 1969, 1970, and 1974, respectively. Since 1974 he has been a member of the Ohio State University, Department of Electrical Engineering, where he is currently an Associate Professor. His primary research interest is in the development of method of moments techniques for the analysis of general antenna or scattering problems, and has published over 40 journal articles in this area. Dr. Newman is a Fellow of the IEEE, a member of Commission B of URSI, and a member of the ELectromagnetics Institute. He is a recipient of the 1986 College of Engineering Research Award, and is a past chairman of the Columbus sections of the IEEE Antennas and Propagation and Microwave Theory and Techniques Societies.

High-Frequency Techniques for Antenna Analysis

PRABHAKAR H. PATHAK, FELLOW, IEEE

Invited Paper

A summary of various high-frequency techniques is presented for analyzing the electromagnetic (EM) radiation from antennas in the presence of their host environment. These techniques not only provide physical insight into antenna radiation mechanisms, but they are found to be highly efficient and accurate for treating a variety of practical antenna configurations. Examples to which these techniques have been applied include open-ended waveguide antennas, horn and reflector antennas, antennas on aircraft and spacecraft, etc. The accuracy of these techniques is established via numerical results which are compared with those based on other independent methods or with measurements. Furthermore, these high-frequency methods can be combined with other techniques, through a hybrid scheme, to solve an even greater class of problems than those which can be solved in an efficient and tractable manner by any one technique alone.

I. Introduction

A summary of some high-frequency techniques is presented for efficiently and accurately analyzing the electromagnetic (EM) radiation from antennas in the presence of their host environment. Such high-frequency techniques also provide a physical insight into the antenna radiation mechanisms involved; this property is useful for both analysis and design purposes.

At sufficiently high frequencies (or short wavelengths), EM wave radiation, propagation, scattering and diffraction, exhibit a highly localized behavior. Such a local description of high-frequency EM waves is given in terms of rays and their associated fields. Thus the total high-frequency field at an observation point is given by the superposition of the fields of all the rays that arrive there, such as via a direct (incident) ray path from the primary antenna excitation (source), and via rays which experience reflection and diffraction from generally different but highly localized regions or "flashpoints" on the antenna and its host structure, as shown for example in Fig. 1. In particular, the incident and reflected rays obey Fermat's principle, and are associated with the usual geometrical optics (GO) incident and reflected fields. In the case of penetrable objects, there also exist GO transmitted rays. On the other hand, the diffracted rays are generally found to originate from geometrical and electrical discontinuities, and from points of grazing incidence on smooth convex portions of the radiating object. The existence of these types of diffracted rays has been postulated by Keller, via an extension of Fermat's principle, in his development of the geometrical theory of diffraction (GTD) [1]; this ray method will be summarized later in more detail. Such a rather simplified and physically appealing picture for the transport of high-frequency EM energy, locally along incident, reflected, and diffracted rays, is in sharp contrast to the description of EM wave radiation at low frequencies that is generally given in terms of the radiation integral on the currents induced globally over the antenna and its entire host structure by the primary antenna excitation. At lower frequencies, one can either employ numerical methods (e.g., moment method, conjugate gradient method, etc.) to solve integral equations for these induced currents, or numerically solve (using finite element or finite difference schemes) the partial differential equations governing the total field behavior. One could also employ a numerical modal (eigenfunction) matching technique for obtaining the relevant field solutions. However, at moderate to high frequencies all of these numerical techniques [64] become very poorly convergent and inefficient because numerical solutions are generally based on exact formulations that must satisfy field self-consistency in a global sense, i.e., over the entire radiating object, rather than requiring a knowledge of the fields in a local sense as done in the high-frequency approximations. It therefore becomes necessary to employ high-frequency techniques for analyzing electrically large radiating objects in a tractable fashion.

One could demonstrate the local nature of high-frequency radiation if one begins by considering the radiation integral over the spatial current distribution induced on a radiating object by the primary excitation. At high frequencies, the

Manuscript received January 5, 1991; revised June 15, 1991. This paper was supported in part by the Joint Services Electronics program under Contract N00014-88-K-004 with the Ohio State University during the preparation of this paper.

The author is with the Department of Electrical Engineering, ElectroScience Laboratory, Ohio State University, Columbus, OH 43212.

IEEE Log Number 9105507.

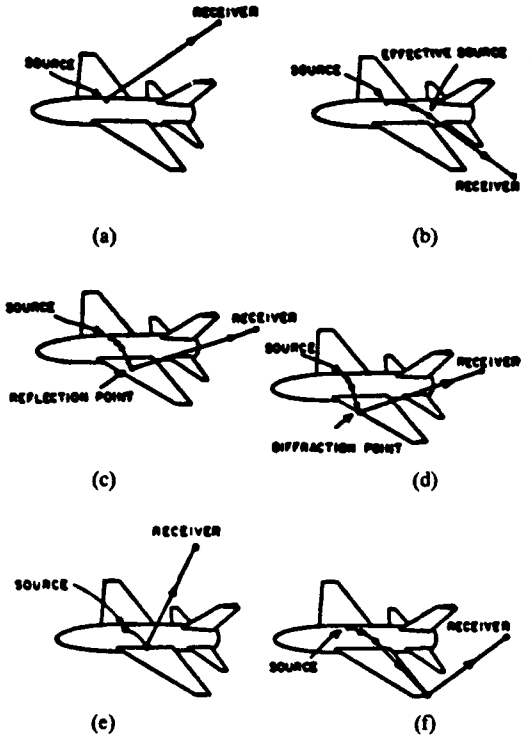


Fig. 1. Rays launched from an infinitesimal antenna element on an aircraft.

phase of the corresponding integrand oscillates rapidly and produces a destructive interference, or cancellation, between the various spherical wave contributions which arrive to a given observation point from the region of integration over the radiating object that excludes any stationary phase points in the integrand. The dominant contribution to the radiation integral then arises from the constructive interference between the spherical waves which emanate from the local neighborhood of any stationary phase points in the region of integration, and also from the end or boundary points of the integral, etc. A similar situation occurs when employing a spectral rather than the spatial representation of the radiation integral; in this case the dominant contribution to the spectral integral for the high-frequency case again arises from the local neighborhood of certain critical points in the spectral integrand, such as saddle points, poles, etc., while a destructive interference generally exists between contributions from the remaining portion of the spectrum. The critical points within the spatial representation of the radiation integral physically correspond to the "flashpoints" or points of reflection, transmission and diffraction on the radiating object. On the other hand, the critical points within the spectral representation for the radiation integral correspond to specific directions, or rays, along which the high-frequency field propagates to the observer. Furthermore, these rays originate from the flashpoints alluded to earlier; consequently, both the spatial and spectral forms of the radiation integral yield the same local picture for the radiation of high-frequency fields. Indeed, a critical point within the radiation integrand of either the spatial or the spectral type leads to the description of a particular ray mechanism (e.g., ray reflection, ray diffraction, etc.) thereby analytically demonstrating the principal of localization of high-frequency fields. Such an evaluation of the radiation integrals in terms of a superposition of the contributions from just the isolated critical points in the integrand constitutes an asymptotic high-frequency approximation for the integrals. Typically, the asymptotic evaluation is performed with respect to a large parameter, e.g., the product of the wavenumber ($2\pi/\lambda$, where λ = wavelength) and some characteristic distance, and the asymptotic approximation becomes increasingly accurate with increase in the large parameter.

It can be verified from an asymptotic evaluation of the radiation integrals, as discussed above, that the ray fields exhibit a "local plane wave" behavior; i.e., the rays are perpendicular to the wavefront (or equiphase) surface in an isotropic medium as shown in Fig. 2; in particular, the wavefront is locally plane in this high-frequency approximation, and the ray field is polarized transverse to the ray. The rays are straight lines in a homogeneous medium. The concept of wavefronts and rays is not new; indeed, it has been central to the development of classical geometrical optics (GO). One recalls that GO includes only the incident, reflected and transmitted ray fields. Limiting the present discussion for the sake of convenience to antennas and their host structures that are impenetrable, it then follows that only the GO incident (or direct) ray from the primary excitation and the GO reflected rays can exist in this case. An example of this situation is shown in Fig. 3 where the line source excites an impenetrable structure; the GO incident and reflected rays exist only in certain portions of the space surrounding this structure. The incident rays (directly radiated from the line source) do not exist beyond the edge induced incident shadow boundary (ISB) and the smooth surface induced surface shadow boundary (SSB), respectively. Also, the reflected rays disappear beyond the edge induced reflection shadow boundary (RSB). Therefore, GO fails to predict a nonzero field within the shadow regions of the incident and reflected rays where such rays cease to exist, and consequently GO cannot describe the diffraction effects behind an impenetrable structure; this may be visualized in Fig. 3. The failure of GO in geometric shadow regions, where the source and its image are not directly visible, was overcome by Keller's GTD [1]. The existence of these diffracted rays in the GTD can be readily verified via the asymptotic reduction of the radiation integrals pertaining to various canonical diffraction problems. Thus according to GTD, the field at the edge Q_E, which is incident from the line source at Q', gives rise to edge diffracted rays emanating from Q_E as in Fig. 4. Likewise the incident ray from Q' which grazes the surface at Q_S launches a surface ray which propagates around the smooth convex boundary transporting energy into the shadow region. Surface diffracted rays are shed along the forward tangent to the surface rays as shown in Fig. 4. The field at P_1 in Fig. 3 consists of simply the GO incident and reflected fields, whereas according to GTD

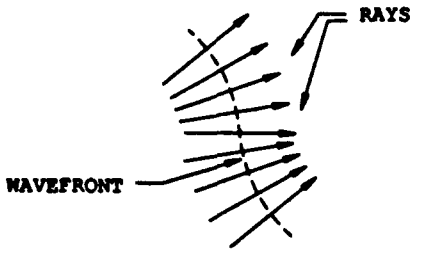


Fig. 2. Wavefront surface and associated family of rays.

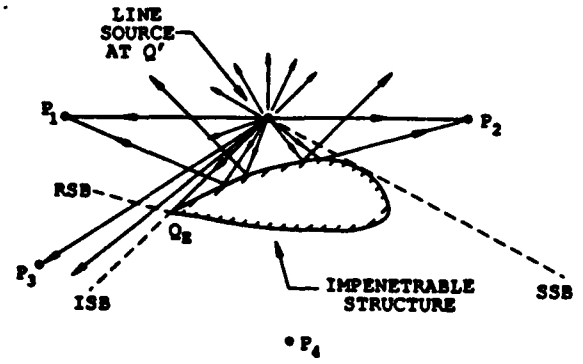


Fig. 3. Geometrical optics (GO) incident and reflected rays produced by a line source radiating in the presence of an impenetrable structure.

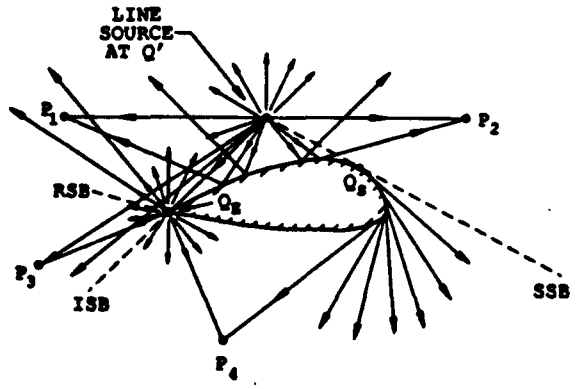


Fig. 4. Edge diffracted rays originating from the edge Q_E and rays diffracted tangentially from the surface ray excited at the point of grazing incidence Q_S.

the field at P_1 in Fig. 4 also contains the additional edge diffracted ray field. The field at P_3 in Fig. 3 is due only to the GO incident field, but GTD again requires that the edge diffracted field be included at P_3 as in Fig. 4. The field at P_4 in Fig. 3 vanishes as predicted by GO; in contrast, the GTD predicts a nonzero field at P_4 which is a superposition of the edge and surface diffracted ray fields as in Fig. 4.

The GTD field is clearly a superposition of GO and diffracted ray fields. Just as the initial amplitudes of the GO reflected and transmitted ray fields are given in terms of the reflection and transmission coefficients, the initial value of a diffracted ray field is likewise given in terms of a diffraction coefficient. The relevant diffraction coefficients may be deduced from asymptotic solutions to simpler canonical problems that model the geometrical and electrical properties of the original problem in the local neighborhood of the point of diffraction. As a result of the extended Fermat's principle, the rays diffracted by an edge lie on a cone about the edge with the cone half angle equal to the angle that the incident ray makes with the edge tangent at the point of diffraction as in Fig. 9. In the case of a two-dimensional problem, the cone of diffracted rays collapses to a disk as in Fig. 4. Furthermore, the surface ray initiated at Q_S in Fig. 4 follows a geodesic path on the convex boundary; also once launched, the surface ray field attenuates as it propagates, because energy is continually shed via rays diffracted tangentially from the surface ray.

Away from the point of diffraction, the GTD diffracted ray field behaves just like a GO ray field. However, such a purely ray optical field description of the GTD fails within the transition regions adjacent to the shadow boundaries (e.g., ISB, RSB, and SSB in Fig. 3) where the GTD diffracted fields generally become singular. The angular extent of the transition region varies inversely with frequency and it also depends on some characteristic distances as will be discussed briefly in Section II. Such a transition region may be viewed as one through which the GTD field changes its ray optical behavior, e.g., as from an incident ray optical type to a diffracted ray optical type across an ISB. This failure of the GTD within the shadow boundary transition regions can be patched up via uniform versions of the GTD such as the uniform geometrical theory of diffraction (UTD) [2]–[4] and the uniform asymptotic theory (UAT) [5]. Additional references dealing with the GTD/UTD/UAT may be found in [6]–[11]. The UTD will be used in this paper as it has been developed for a variety of canonical shapes, whereas the UAT has been developed only for an edge at the present time.

It was indicated earlier that the asymptotic evaluation of radiation integrals gives rise to a total high-frequency field in terms of a superposition of the contributions from certain isolated or critical points in the integrand (which can be seen to correspond to the field of GTD rays). Some of these critical points can come close together when the observation point lies within the shadow boundary transition regions, and even coalesce for an observer on the shadow boundary itself. This leads to a "coupling between the critical points" and the asymptotics must then be modified; i.e., it must be performed via a uniform procedure which accounts for this coupling and thus forms the basis of the UTD. In the UTD, the GTD solution is modified through the use of uniform asymptotic procedures which systematically introduce additional factors, referred to as the UTD transition functions. These UTD transition functions compensate the GTD singularities at shadow boundaries and keep the total high-frequency field bounded, and continuous, across these boundaries, thus keeping the field valid within the transition regions. Furthermore, outside the shadow boundary transition layers, the UTD automatically reduces to the GTD. These transition functions are special functions characteristic of the diffraction process; e.g., in the case of

edge diffraction they involve Fresnel integrals, whereas in the case of convex surface diffraction they involve Fock functions [12] which contain integrals of Airy functions. The latter functions are named after V. A. Fock who contributed significantly to the analysis of wave diffraction by smooth convex boundaries.

It thus follows from the preceding paragraph that it is the UTD and not the GTD which must be used in practical applications to obtain continuous total (high-frequency) fields (around the radiating object). Besides the singularities of the GTD at the GO shadow boundaries discussed above, the GTD and its uniform versions such as the UTD, UAT, etc., exhibit singularities at the caustics of GO and diffracted rays. Ray caustics occur whenever a family of rays (i.e., ray congruences) merge or intersect; examples of ray caustics are shown in Figs. 5 and 6. In particular, the diffracted ray caustic at P in Fig. 5 is produced on the axis of a symmetric parabolic reflector illuminated by a feed at the focus. The smooth caustic of reflected rays in Fig. 6 is produced by a shaped subreflector which is a surface of revolution; it is illuminated by a feed antenna located on the subreflector axis. This subreflector surface exhibits an inflection point along its generator giving rise to the caustic. Such a smooth caustic can also be produced by a concave reflector surface. A curved edge can likewise generate a smooth caustic of diffracted rays. Ray caustics can become problematic in the GTD/UTD/UAT computations only if they occur in real space (exterior to the antenna and its host structure); otherwise, they are of little concern whenever they occur in virtual space e.g., within the scatterer or the antenna host structure, unless the transition region adjacent to the virtual caustic emerges into external space where a field or observation point may be located. The failure of the GTD/UTD at GO or diffracted ray caustics and their associated transition regions can be patched up through a uniformizing procedure which again introduces special functions (or caustic transition functions) to correct the pure ray solution. For a smooth caustic as in Fig. 6, the special transition function involves the Airy function and its derivative [13], [14]; if the caustic curve has a cusp then one obtains Pearcey functions (related to the parabolic cylinder functions) [15]. If either the smooth or the cusped caustic terminates, as might happen when the reflecting surface terminates at an edge, then one requires incomplete Airy functions or incomplete Pearcey functions, respectively [15], to evaluate fields near the caustic termination. These special functions (or transition functions) reflect the coupling of the pertinent critical points in the asymptotic evaluation of the radiation integral as discussed earlier.

A procedure which can treat more general diffracted ray caustic effects is based on the equivalent current method (ECM) [16], [17]. The ECM while primarily useful for handling caustics of diffracted rays can in some special cases also be employed to handle caustics of reflected rays. In general, the ECM, which corrects for the singularities of the fields at diffracted ray caustics that lie outside the ISB, RSB and SSB transition regions (where UTD

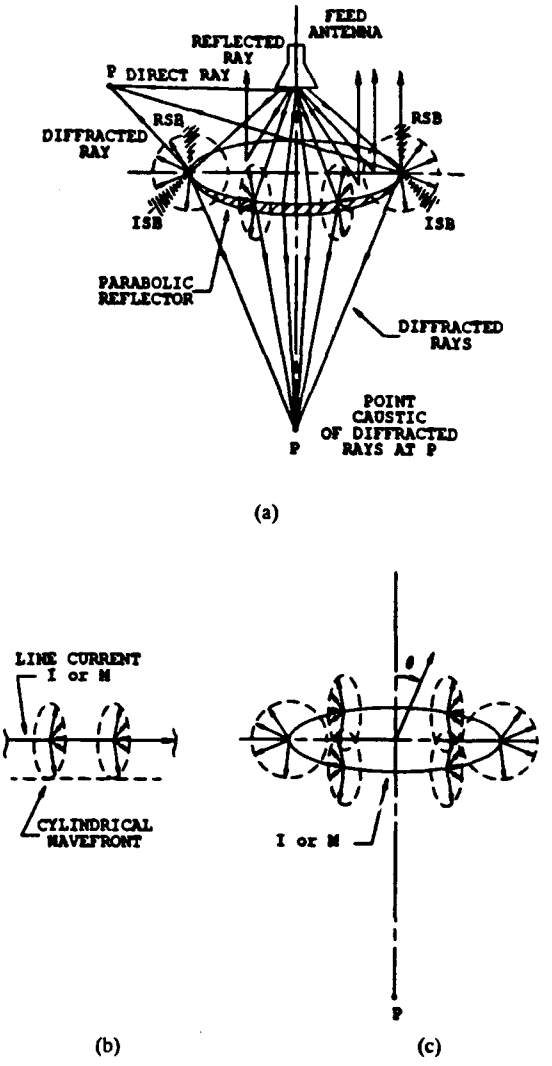


Fig. 5. Point caustic of edge diffracted rays on the axis of a symmetrically fed parabolic reflector.

reduces to GTD) describes the fields within such caustic transition regions in terms of a radiation integral over a set of equivalent *line* currents that replace the original geometry and its illumination; these equivalent currents are found from the GTD diffraction coefficients which are not singular in caustic directions even though the actual GTD ray field is singular there [4]. The fields in the diffracted ray caustic region of Fig. 5 can be treated by ECM. This ECM procedure, if formulated properly, is a uniform procedure in that away from the caustic transition region, the ECM radiation integral reduces asymptotically to the GTD. In a few cases, the ECM radiation integral reduces to a closed form result, or it can be expressed in terms of special functions (e.g., Airy or Pearcey functions) alluded to above; however, in general the integral must be evaluated numerically.

The diffraction effects within the GO shadow boundary transition regions are generally not localized to just the edges or points of diffraction because they are then coupled to the GO effects on the reflecting surface, and in order

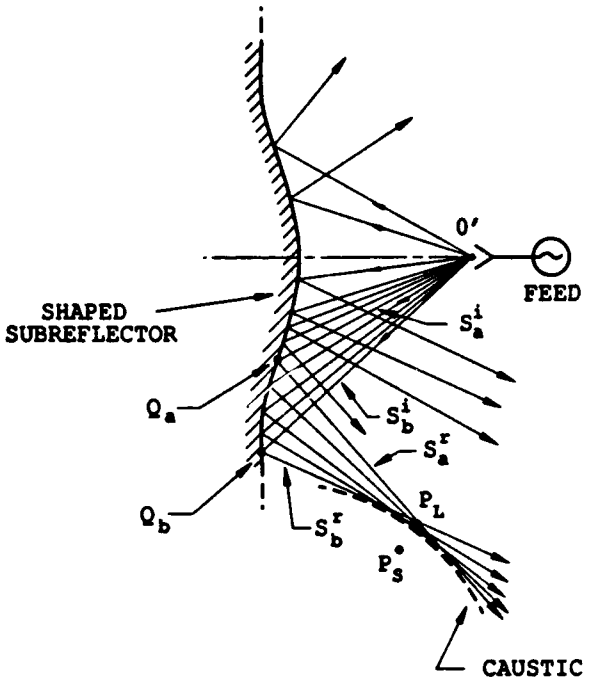


Fig. 6. Ray caustic generated by a shaped subreflector illuminated by a feed antenna.

to correct for the singularities of GTD ray caustics in directions where the associated caustic transition regions overlap with the GO shadow boundary transition regions, one must therefore resort to a *surface* integral representation and not the *line* integral approximations of ECM. The physical optics (PO) surface integral approach [18] and its modifications based on Ufimtsev's physical theory of diffraction (PTD) [19], [20] for edged bodies as discussed in Section II-C, and Fock's theory for curved bodies [12], become useful for treating the fields within the overlap of diffracted and/or GO ray caustic and GO ray shadow boundary transition regions. An example of the overlap of the caustic and GO shadow boundary transition regions is again provided by Fig. 5, where the RSB coincides with the forward axial caustic of the edge diffracted rays in the far zone of the reflector. Furthermore, there are also an infinite number of rays reflected from the parabolic surface which contribute to the far field in this forward axial direction. The forward axial direction of the parabolic reflector is therefore also a caustic of reflected rays in the far zone, in addition to being a caustic of the edge diffracted rays. While the PO method in itself gives quite accurate results for caustic fields in the region of the overlap of the GO shadow boundary and caustic transition regions, the PTD which provides a correction to PO can yield more accurate results outside the caustic region. *It is noted that the PTD is a superposition of PO and the correction to PO as specified by Ufimtsev.*

In general, the integrals in the PTD approach must be evaluated numerically; only in special cases can they be evaluated in closed form. In situations where the integrals in the PTD can be evaluated asymptotically, they recover the leading terms of the GTD fields. Furthermore, if these PTD integrals can be evaluated asymptotically in a uniform fashion, then PTD can be shown to recover the UTD. The intimate connection between the PTD and the GTD/UTD indicated above allows one to view PTD as an integral version of the GTD/UTD which is valid even in regions of ray caustics, and also in regions of overlap of caustic and GO shadow boundary transition regions where GTD/UTD fails. However, PTD requires an integration whereas UTD does not. Therefore, it appears to be far more efficient to employ UTD everywhere except at ray caustics, and in the overlap of caustic and GO shadow boundary transition regions, where the more general PTD integrals may be used to patch up the UTD. While the PTD is a high-frequency technique in its own right like the GTD/UTD, it has been developed only for edges as indicated previously, whereas the UTD can also handle surface diffraction and other types of diffraction mechanisms. Furthermore, the PTD cannot account for multiple diffraction effects as easily as the GTD/UTD; such higher order multiple wave interactions can become important if the scattering/diffraction centers come close together on a radiating object (e.g., if a pair of interacting edges come close together). Nevertheless, the PTD has been often used for predicting the dominant contribution to the radar cross section (RCS) of complex targets (e.g., aircraft, missiles, etc.).

The above-mentioned high-frequency techniques based on the GTD/UTD, ECM, and PTD will be applied to some illustrative antenna examples in the next section. An $e^{+j\omega t}$ time convention for the sources and fields will be assumed and suppressed in the following work. Also, k is assumed to be the wavenumber in the isotropic homogeneous medium external to the antenna and its host structure ($k = 2\pi/\lambda$; λ = wavelength in the external medium).

II. Summary of High Frequency Techniques with Specific Antenna Applications

The high-frequency techniques such as the GTD/UTD, ECM, and PTD, which have been briefly discussed above in Section I, are reviewed in slightly more detail in this section and results based on these techniques are illustrated for dealing with some antenna geometries of interest. The main focus will be on GTD/UTD-based applications; these will be discussed first. The applications based on the ECM and the PTD will be illustrated next only from the perspective of patching up GTD/UTD in those few special situations where the latter fails as discussed previously, such as in regions of ray caustics, and where there is a confluence of caustic and GO shadow boundary transition regions.

A. GTD and Its Uniform Version, UTD

As discussed in Section I, the GTD/UTD is a ray technique. Therefore, it would be worthwhile to firstly develop a general expression for the ray optical field. While there are several procedures, involving either the relevant asymptotic approximations of radiation integrals pertaining to

certain canonical radiation problems, or the asymptotic (Luneberg–Kline) series expansion of the wave equation, to arrive at a ray optical field expression, the simpler and less rigorous approach based on geometrical considerations is employed here. In particular, since energy in the high-frequency EM field is assumed to be transported along rays, it follows from geometrical considerations that power must be conserved in each narrow tube of rays (or a ray pencil) in a lossless medium. Thus consider any given (central or axial) ray OP in a ray tube as shown in Fig. 7. Let the principal wavefront radii of curvature at O be ρ_1 and ρ_2, respectively; the corresponding principal wavefront radii of curvature of that ray at P are $(\rho_1 + s)$ and $(\rho_2 + s)$, where $|\overline{OP}| = s$. Let the electric field intensity at O and P be $\overline{E}(O)$ and $\overline{E}(P)$, respectively; thus, the power crossing the area dA_o is given by $(1/Z_o)|\overline{E}(O)|^2 dA_o$ where $dA_o \approx |(\rho_1 d\psi_1)(\rho_2 d\psi_2)|$ and where Z_o is the plane wave impedance in the medium; likewise, the power crossing dA_p is $(1/Z_o)|\overline{E}(P)|^2 dA_p$ where $dA_p \approx |[(\rho_1 + s)d\psi_1][(\rho_2 + s)d\psi_2]|$. Conservation of power in the ray tube requires $(1/Z_o)|\overline{E}(P)|^2 dA_p = (1/Z_o)|\overline{E}(O)|^2 dA_o$; i.e.,

$$|\overline{E}(P)| = |\overline{E}(O)| \left| \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + s)(\rho_2 + s)}} \right|. \tag{1a}$$

Incorporating the local plane wave polarization and phase heuristically into (1a) yields the rule for continuation of the field $\overline{E}(O)$ at O to the field $\overline{E}(P)$ at P along the ray $\overline{OP}$ as

$$\overline{E}(P) \sim \overline{E}(O) \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + s)(\rho_2 + s)}}\, e^{-jks}. \tag{1b}$$

The field in (1b), which is referred to as an arbitrary ray optical field (where ρ_1 and ρ_2 are arbitrary), can be shown to reduce to a plane wave (if $[\rho_1, \rho_2] \to \infty$), cylindrical or conical wave (if ρ_1 or $\rho_2 \to \infty$), and a spherical wave (if $\rho_1 = \rho_2 =$ finite value), respectively. Thus the latter more familiar wave types are all special cases of a ray optical field whose general form is (1b). One notes that the ray congruences at 1–2 and 3–4 form a ray caustic (or centers of radii of curvature ρ_1 and ρ_2, respectively of the wavefront) in Fig. 7. The ρ_1 and ρ_2 are positive if the ray caustics at 1–2 and 3–4 occur before reaching the reference point O along the ray direction $\hat{s}$ in Fig. 7; otherwise, they are negative. The positive branch of the square root is chosen in (1b); hence, if $\rho_{1,2} < 0$ and $s > -|\rho_2|$ or $s > -|\rho_1|$, then a caustic is crossed at 1–2 or 3–4, respectively, and $(\rho_2 + s)$ or $(\rho_1 + s)$ changes sign so that a phase jump of $\pi/2$ due to caustic traversal needs to be included in (1b), because

$$\sqrt{\frac{\rho_i}{\rho_i + s}} = \left| \sqrt{\frac{\rho_i}{\rho_i + s}} \right| e^{j\pi/2},$$
$$\text{if } \rho_i = -|\rho_i| \text{ and } s > -|\rho_i|, \text{ for } i = 1, 2.$$

Furthermore, the magnetic field $\overline{H}(P)$ at P is found from $\overline{E}(P)$ in (1b) via the local plane wave condition along a ray; namely,

$$\overline{H}(P) = Y_o \hat{s} \times \overline{E}(P) \tag{2}$$

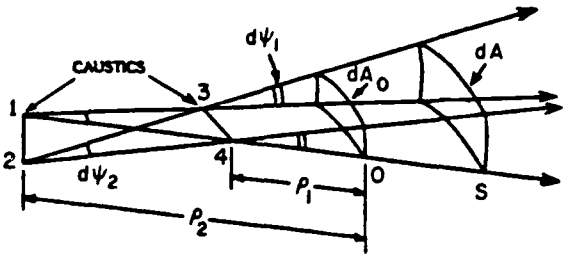


Fig. 7. Ray tube (ray pencil).

where $Y_o = 1/Z_o$, and Z_o as before is the local plane wave impedance of the medium in which the ray propagates.

The quantity $\overline{E}(P)$ in (1b) may represent a field which is associated with either an incident ray, or with rays that are reflected or transmitted at an interface between two media, or with diffracted rays. The initial ray amplitudes at the points of reflection and transmission can be found by enforcing the EM boundary conditions at the interface; these conditions also lead to Snell's laws of reflection and transmission which are consistent with Fermat's principle and which could in fact have been derived from it. The initial value of the diffracted ray field is given in terms of the diffraction coefficient and the diffracted ray path obeys the extended Fermat's principle. Thus in general, (1b) can be written as

$$\overline{E}^p(P) \sim \overline{E}^p(O_p) \sqrt{\frac{\rho_1^p}{\rho_1^p + s^p} \cdot \frac{\rho_2^p}{\rho_2^p + s^p}}\, e^{-jks^p} \tag{3a}$$

$$\text{with } p \to i, r \text{ or } d \tag{3b}$$

where the superscript or subscript, p, refers to the incident (i), reflected (r) or diffracted (d) ray fields.

1) Incident GO Ray Field: Letting $p = i$ in (3a) and (3b) allows one to write the GO incident ray field as

$$\overline{E}^i(P) = \overline{E}^i(O_i) \sqrt{\frac{\rho_1^i}{\rho_1^i + s^i} \cdot \frac{\rho_2^i}{\rho_2^i + s^i}}\, e^{-jks^i} U_i \tag{4a}$$

where U_i is unity in the region where the GO incident ray field exists and is zero otherwise. The incident principal wavefront radii of curvature ρ_1^i and ρ_2^i are measured from the reference point O_i along the incident ray to P. It is noted that $s^i = |\overline{O_i P}|$. In the two-dimensional case $\rho_2^i \to \infty$ and (4a) becomes

$$\overline{E}^i(P) = \overline{E}^i(O_i) \sqrt{\frac{\rho_1^i}{\rho_1^i + s^i}}\, e^{-jks^i} U_i\,. \tag{4b}$$

2) Reflected GO Ray Field: An expression for the reflected ray field can be obtained by letting $p = r$ in (3a) and (3b), and by letting the point $O_p = O_r$ move to the point of reflection Q_R, then (3a) and (3b) become

$$\overline{E}^r(P) \sim \overline{E}^r(Q_R) \sqrt{\frac{\rho_1^r}{\rho_1^r + s^r} \cdot \frac{\rho_2^r}{\rho_2^r + s^r}}\, e^{-jks^r} U_r. \tag{5a}$$

where the step function U_r is unity in the region where the reflected ray field exists and is zero otherwise. The reflected

field $\overline{E}^r(Q_R)$ at the point of reflection Q_R can be related to the incident field $\overline{E}^i(Q_R)$ at Q_R by the dyadic surface reflection coefficient $\overline{\overline{R}}$ as follows:

$$\overline{E}^r(Q_R) = \overline{E}^i(Q_R) \cdot \overline{\overline{R}}. \tag{5b}$$

Incorporating (5b) into (5a) yields the required expression for the reflected field at P due to the field incident on the surface at the point of reflection Q_R as

$$\overline{E}^r(P) \sim \overline{E}^i(Q_R) \cdot \overline{\overline{R}} \sqrt{\frac{\rho_1^r}{\rho_1^r + s^r} \cdot \frac{\rho_2^r}{\rho_2^r + s^r}} e^{-jks^r} U_r. \tag{5c}$$

The reflected wavefront radii of curvature $\rho_{1,2}^r$ are shown in Fig. 8 and they may be calculated via the expressions given in [2]. The dyadic reflection coefficient $\overline{\overline{R}}$ may be found by approximating the original surface locally by a plane tangent to that surface at the point of reflection Q_R when it is illuminated by an EM plane wave, and by enforcing the EM boundary conditions at Q_R. It is convenient to express the incident and reflected fields in terms of the unit vectors fixed in the incident and reflected rays as in Fig. 8. Let $\hat{e}_\parallel^i$ and $\hat{e}_\parallel^r$ be unit vectors fixed in the plane of incidence containing the unit normal vector $\hat{n}$ to the surface at Q_R and the incident ray direction $\hat{s}^i$ at Q_R, and let these vectors also be perpendicular to the incident and reflected ray directions $\hat{s}^i$ and $\hat{s}^r$, respectively. Likewise, let $\hat{e}_\perp$ be a unit vector perpendicular to the plane of incidence at Q_R. In these ray fixed unit vectors, $\overline{\overline{R}}$ becomes

$$\overline{\overline{R}} = \hat{e}_\parallel^i \hat{e}_\parallel^r R_h + \hat{e}_\perp \hat{e}_\perp R_s \,. \tag{6}$$

For a perfectly conducting surface, $R_s = -1$ and $R_h = 1$. If the reflecting boundary and illumination becomes two-dimensional, then $\rho_2^r \to \infty$ and (5a) becomes

$$\overline{E}^r(P) \sim \overline{E}^i(Q_r) \cdot \overline{\overline{R}} \sqrt{\frac{\rho_1^r}{\rho_1^r + s^r}} e^{-jks^r}$$

(two-dimensional case) (7)

3) Edge Diffracted Ray Field: Consider an edge diffracted ray field produced by an incident wave which strikes a wedge at Q_E as in Fig. 9. One can obtain the general expression for the edge diffracted ray field once again from (3a) and (3b) by letting $p = d$ so that

$$\overline{E}^d(P) \sim \overline{E}^d(O_d) \sqrt{\frac{\rho_1^d}{\rho_1^d + s^d} \cdot \frac{\rho_2^d}{\rho_2^d + s^d}} e^{-jks^d}. \tag{8a}$$

It is useful to move the reference point O_d along the edge diffracted ray at P to the point Q_E at the point of diffraction on the edge; thus $\rho_1^d \to 0$ as can be seen from Fig. 9. Even though $\rho_1^d \to 0$ in (8a) one can show that:

$$\lim_{\substack{\rho_1^d \to 0 \\ O_d \to Q_E}} \left[\overline{E}^d(O_d) \sqrt{\rho_1^d} \right] = \overline{E}^i(Q_E) \cdot \overline{\overline{D}}_e \,. \tag{8b}$$

where $\overline{E}^i(Q_E)$ is the field of the ray incident at Q_E and $\overline{\overline{D}}_e$ is the dyadic edge diffraction coefficient. Incorporating

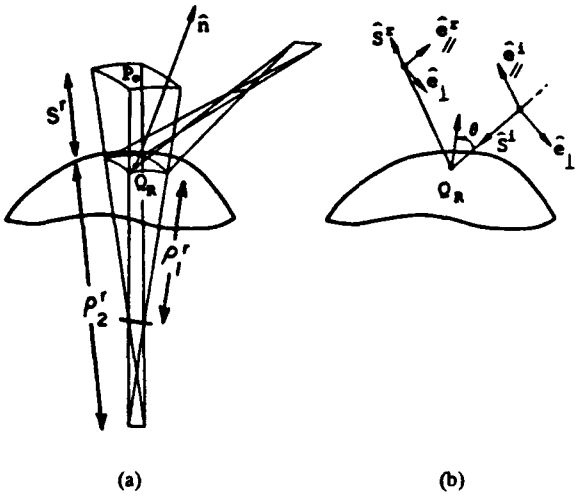


Fig. 8. Reflected wavefront curvatures and unit vectors associated with the reflection problem.

(8b) into (8a) with $\rho_1^d \to \rho^d$ as $O_d \to Q_E$ gives

$$\overline{E}^d(P) \sim \overline{E}^i(Q_E) \cdot \overline{\overline{D}}_e \sqrt{\frac{1}{s^d} \cdot \frac{\rho^d}{\rho^d + s^d}} \, e^{-jks^d} \,. \tag{8c}$$

It is convenient to express the incident field $\overline{E}^i(Q_E)$ and the diffracted field $\overline{E}^d(P)$ in terms of unit vectors fixed in the incident and diffracted rays as shown in Fig. 9. The edge tangent $\hat{e}$ at Q_E and the incident ray direction $\hat{s}'$ form the edge fixed plane of incidence. The unit vectors $\hat{\beta}_o'$ and $\hat{\phi}'$ are parallel and perpendicular, respectively, to the edge fixed plane of incidence. Likewise $\hat{\beta}_o$ and $\hat{\phi}$ are parallel and perpendicular, respectively, to the edge fixed plane of diffraction formed by $\hat{e}$ and the diffracted ray direction $\hat{s}^d$. Furthermore, $\hat{\beta}_o' = \hat{s}' \times \hat{\phi}'$ and $\hat{\beta}_o = \hat{s}^d \times \hat{\phi}$. Thus $\overline{E}^i = \hat{\beta}_o'(\hat{\beta}_o' \cdot \overline{E}^i) + \hat{\phi}'(\hat{\phi}' \cdot \overline{E}^i)$ and $\overline{E}^d = \hat{\beta}_o(\hat{\beta}_o \cdot \overline{E}^d) + \hat{\phi}(\hat{\phi} \cdot \overline{E}^d)$, so that $\overline{\overline{D}}_e$ is given by [2]:

$$\overline{\overline{D}}_e = -\hat{\beta}_o' \hat{\beta}_o D_{es}(\phi, \phi'; \beta_o) - \hat{\phi}' \hat{\phi} D_{eh}(\phi, \phi'; \beta_o) \,. \tag{9}$$

The scalar UTD edge diffraction coefficients D_{es} and D_{eh} contained in the dyadic UTD edge diffraction coefficient $\overline{\overline{D}}_e$ are obtained from a uniform asymptotic solution to the canonical problem of the diffraction of plane, cylindrical, conical and spherical incident waves by a perfectly conducting wedge [2] and they contain a sum of four simple terms, each of which is a product of a cotangent function (involving ϕ, ϕ', and β_o) and a transition function F containing a Fresnel integral, where

$$F(x) = 2j\sqrt{x} e^{jx} \int_{\sqrt{x}}^{\infty} d\tau \, e^{-j\tau^2} \,. \tag{10}$$

The argument of the F functions depend on the incident, reflected and diffracted wavefront curvatures, and they are defined in [2]; the F function, which is well tabulated, is responsible for keeping the $\overline{E}^d$ bounded at the GO shadow boundaries where GTD predicts a singularity. Outside the GO shadow boundary transition regions, the F function

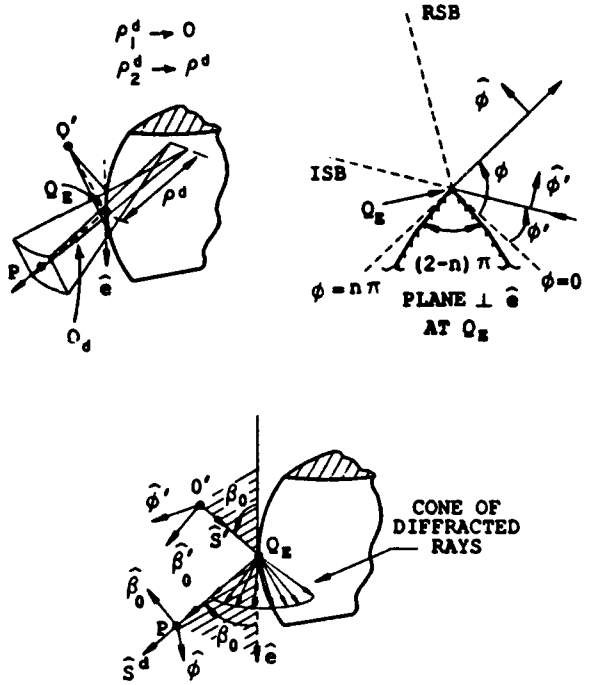


Fig. 9. Wedge diffraction geometry.

becomes unity and the UTD result then reduces automatically to GTD. In the two-dimensional case, $\rho^d \to \infty$ in (8c) so that $\overline{E}^d(P) = \overline{E}^i(Q_E)\cdot\overline{\overline{D}}_e\sqrt{(1/s^d)}e^{-jks^d}$ for two-dimensional edge configurations. Some examples which can be analyzed using these UTD edge diffraction concepts are indicated below.

Consider the symmetric parabolic reflector antenna with a feed whose phase center is at the focus of the parabola as shown in Fig. 10. The UTD electric field at P_{01} in the near zone of this reflector as shown in Fig. 10 is then given by

$$\overline{E}(P_{01}) \sim \overline{E}^i(P_{01}) + \overline{E}_1^d(P_{01}) + \overline{E}_2^d(P_{01}) \qquad (11)$$

where the field $\overline{E}^i$ directly radiated by the feed to P_{01} has the form

$$\overline{E}^i(P_{01}) \sim c\overline{f}(\theta,\phi)\frac{e^{-jks^i}}{s^i}U_i \qquad (12)$$

with

$$U_i = \begin{cases} 1, & \text{in region where the feed is directly visible,} \\ 0, & \text{behind the reflector (within } ISB_1 \text{ and } ISB_2\text{)} \\ & \text{where the feed is shadowed.} \end{cases}$$

The quantity c in (12) is a known complex constant, and $\overline{f}(\theta,\phi)$ is the vector radiation pattern of the feed with θ measured with respect to the z axis while ϕ is the azimuthal angle about this axis of symmetry of the paraboloidal reflector antenna; the quantity $\overline{f}$ is also assumed to be known. The field in (12) constitutes a spherical wave from the feed. The fields $\overline{E}_1^d$ and $\overline{E}_2^d$ are diffracted from two distinct points on the edge of the reflector, where the plane containing P_{01} and the reflector axis intersects the edge at Q_1 and Q_2 in accordance with the extended Fermat's

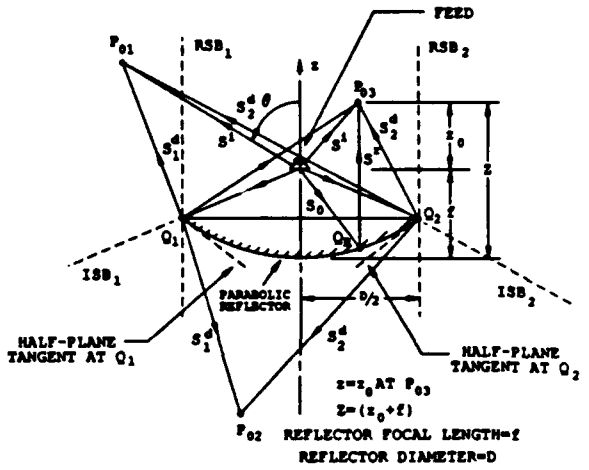


Fig. 10. Rays contributing to the near field of a symmetric parabolic reflector antenna.

principle; these will be described in (16). Next, the field $\overline{E}(P_{02})$ at P_{02} also in the near zone behind the reflector is:

$$\overline{E}(P_{02}) \sim \overline{E}_1^d(P_{02}) + \overline{E}_2^d(P_{02}) \quad . \qquad (13)$$

Likewise, the field $\overline{E}(P_{03})$ at P_{03} in the near zone becomes

$$\overline{E}(P_{03}) \sim \overline{E}^i(P_{03}) + \overline{E}^r(P_{03}) + \overline{E}_1^d(P_{03}) + \overline{E}_2^d(P_{03}) \qquad (14)$$

in which $\overline{E}^r$ is the field at P_{03} that is reflected from Q_R as in Fig. 10; it is given via (5c) by

$$\overline{E}^r(P_{03}) \sim \overline{E}^i(Q_R)\cdot\overline{\overline{R}}\sqrt{\frac{\rho_1^r\rho_2^r}{(\rho_1^r+s^r)(\rho_2^r+s^r)}}e^{-jks^r}U_r \qquad (15)$$

with,

$$U_r = \begin{cases} 1, & \text{within the region containing the } z\text{-axis} \\ & \text{and bounded by } RSB_1 \text{ and } RSB_2 \\ 0, & \text{otherwise.} \end{cases}$$

Also, $\overline{E}^i(Q_R) \sim c\overline{f}(\theta_o,\phi_o)(e^{-jks_o}/s_o)$ in (15) where (θ_o,ϕ_o) are the values of (θ,ϕ) along the direction $\hat{s}_o = \overrightarrow{(OQ_R)}/(|\overrightarrow{OQ_R}|)$. Finally, the two edge diffracted ray fields $\overline{E}_1^d$ and $\overline{E}_2^d$ in (11), (13), and (14) have the general form:

$$\overline{E}_j^d(P_o) \sim \left[\overline{E}^i(Q_j)\cdot\overline{\overline{D}}_e(Q_j) + \left.\frac{\partial\overline{E}^i}{\partial n_j^i}\right|_{Q_j} \cdot\overline{\overline{d}}_e^i(Q_j) + \left.\frac{\partial\overline{E}^r}{\partial n_j^r}\right|_{Q_j}\cdot\overline{\overline{d}}_e^r(Q_j)\right] \cdot\sqrt{\frac{\rho^{dj}}{s_j^d(\rho^{dj}+s_j^d)}}e^{-jks_j^d} \qquad (16)$$

with $j = 1,2$ corresponding to Q_1 and Q_2. While the term involving $\overline{E}^i(Q_E)\cdot\overline{\overline{D}}_e$ in (16) has been introduced in the discussion on the UTD for edge diffraction leading to (8c), the term containing $\overline{\overline{d}}_e^{i,r}$ in (16) is an additional contribution to the UTD edge diffracted field, and it is termed as the slope diffraction contribution [3], [4]. The

slope diffraction contribution becomes important if the incident field $\overline{E}^i$ and/or the reflected field $\overline{E}^r$ exhibits a rapid spatial variation near the point of diffraction Q_E on the edge. For example, if the field $\overline{E}^i$ vanishes at the point of diffraction Q_E, then the field diffracted from Q_E calculated via $\overline{E}^i(Q_E) \cdot \overline{\overline{D}}_e$ would also vanish; however, if $\overline{E}^i(Q_E)$ tends to vanish rapidly at Q_E so that its spatial derivative symbolized here by $\partial \overline{E}^i / \partial n^i$ may be significant, then it could give rise to the slope edge diffraction contribution which must therefore be included for accuracy. In the present application, if the feed pattern $\overline{f}(\theta, \phi)$ is rapidly varying at the edges $Q_{1,2}$ then the slope diffraction term in (20) will be important; otherwise, the slope effects are generally negligible. The results in (11)–(16) have been employed in [21] to obtain the near field radiation from a parabolic reflector antenna in the plane $Z = z_o + f$ in Fig. 11. However, the results in (11) and (13) can also be used in the far zone of the reflector outside the paraxial region. The numerical results in Fig. 11 based on the UTD as obtained in [21] are compared with those based on GO ($= \overline{E}^i + \overline{E}^r \approx \overline{E}^r$ in the forward direction since $|\overline{E}^i| \ll |\overline{E}^r|$ for the feed employed in this example), and with the commonly used but far less efficient aperture integration (AI) technique. It is noted that the GO reflected field $\overline{E}^r$ is discontinuous in Fig. 11 as required by U_+ in (15). Also, the agreement between UTD and the reference solution based on AI is quite good in that figure. Finally, it is noted that, for a small range of angles near the plane of the reflector, one of the edges is always shadowed by the reflector geometry, and this shadow zone is filled by surface rays which are excited on the back (convex) side of the reflector surface via edge diffraction, and these rays then shed energy tangentially as surface diffracted rays. A whispering gallery type field can also be excited on the concave front side of the reflector via edge diffraction. Such edge excited surface diffracted rays [22]–[25], and the diffraction of whispering gallery fields [26] occur in a small angular region and may generally be neglected to first order without incurring serious errors.

4) UTD Corner (Vertex) Diffracted Field: Corners or vertices can occur if an edge is truncated, e.g., as in the case of a plane angular sector, or a finite plate structure for which the edge tangent is discontinuous (to form the corner), or as in the case of a pyramidal structure with planar facets whose edges converge to a point; these specific examples are illustrated in Figs. 12(a) and (b). In addition, corners or tips can also occur in a smooth conical geometry, as shown in Fig. 12(c). When a corner in an impenetrable surface is illuminated by a source, then the incident ray is diffracted in all radial directions from the corner as shown in Fig. 12. The UTD field of these corner diffracted rays illustrated in Fig. 12(a) and (b) keeps the total high-frequency field bounded and continuous across the corner induced shadow boundaries of rays diffracted by the edges, just as the UTD edge diffracted fields keep the total high-frequency field bounded and continuous across the edge induced shadow boundaries of the GO incident and reflected

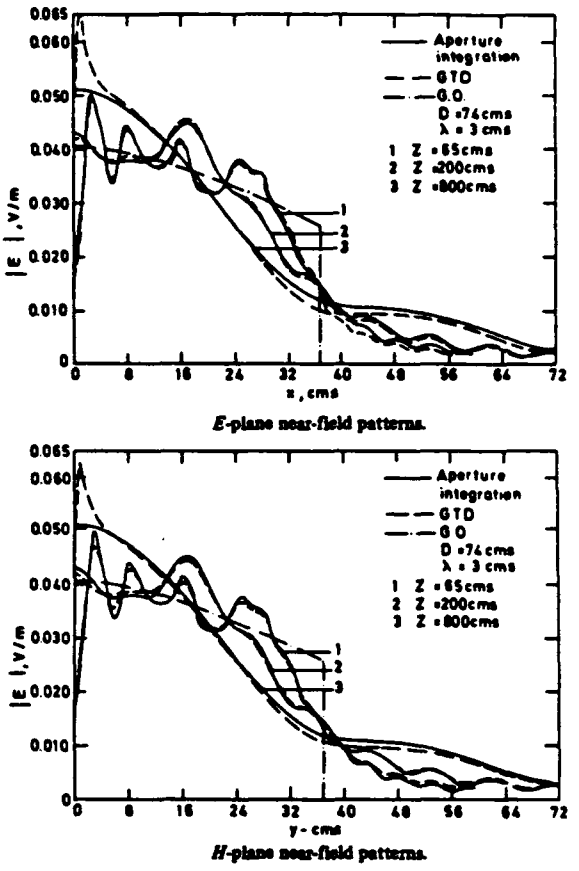


Fig. 11. Near field radiation patterns of a parabolic reflector antenna (from [21]).

rays. The corner induced shadow boundaries of the edge diffracted rays of course occur because such rays cease to exist whenever the edges terminate (at the corner). The general UTD form of the corner diffracted field is given by

$$\overline{E}_c^d(P) \sim \overline{E}^i(Q_C) \cdot \overline{\overline{D}}_c \frac{e^{-jks_c^d}}{s_c^d} \tag{17}$$

$$\overline{H}_c^d(P) = Y_o \hat{s}_c^d \times \overline{E}_c^d(P). \tag{18}$$

Recently, an approximate but useful UTD result for $\overline{\overline{D}}_c$ has been obtained for the case of the diffraction by a corner in a perfectly conducting plane angular sector as shown in Fig. 12(a). The UTD transition function present in $\overline{\overline{D}}_c$, which compensates for the corner induced discontinuity in the edge diffracted fields may be viewed as an integral of a Fresnel integral that can be calculated quite efficiently. The present more rigorously obtained UTD corner diffraction coefficient in [27], [28] constitutes an improvement over a previous one which was constructed heuristically [29]. Figure 13 indicates the far zone radiation pattern of a dipole antenna located near a perfectly conducting rectangular plate, which has been calculated via the new $\overline{\overline{D}}_c$ in [27], [28]; this pattern is seen to compare very well with an independent moment method (MM) solution of an integral equation for the problem.

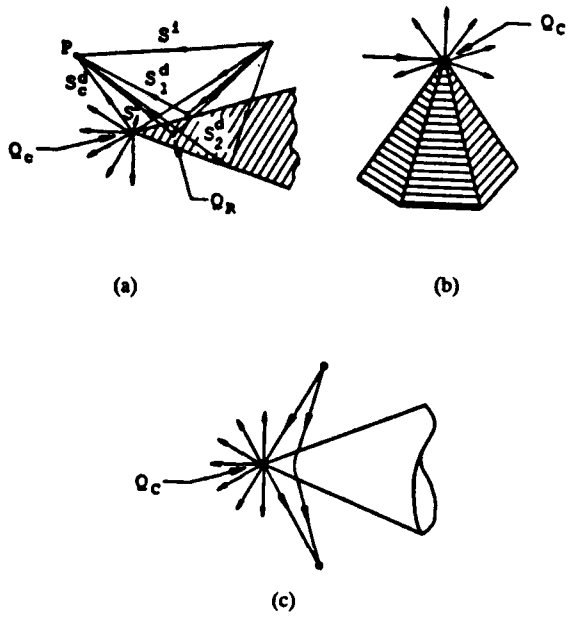


Fig. 12. Examples of corner diffracted rays.

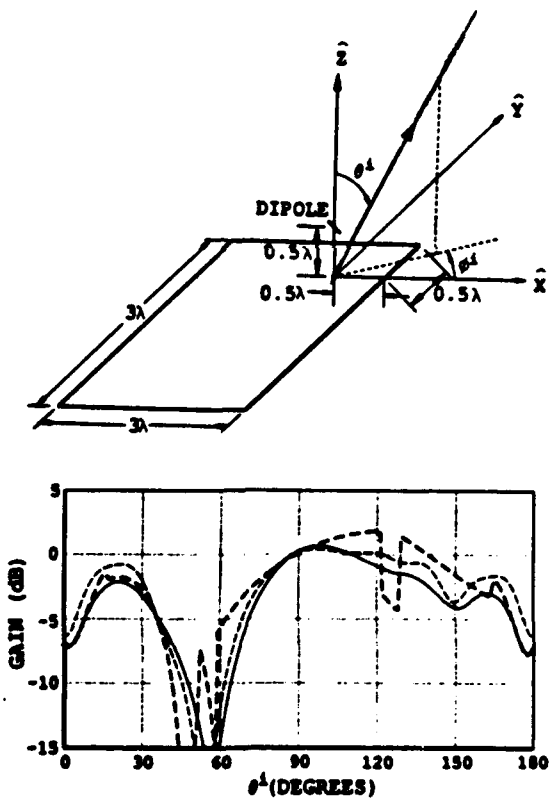


Fig. 13. Radiation pattern of a dipole near a perfectly conducting rectangular plate. (—— UTD ; - - - MM ; – – – UTD without corner diffraction term). Note: $\phi^i = 47.5°$.

For the case in Fig. 12(c), there are no edge diffracted rays because the cone geometry is smooth (except at the tip) and contains no edges; however, there are surface diffracted rays that are launched from the smooth portion of the conical surface. The nature of the surface diffracted rays changes rapidly as the surface rays on the cone approach the cone tip because of the rapid decrease in the surface radius of curvature (except along the cone generator) as the tip is approached. The UTD cone tip diffracted field must then contain information on this change in the nature of the surface ray field near the tip; such a general UTD solution has not yet been developed although some initial attempts have been made [30]. On the other hand, an approximate UTD-type diffraction coefficient based on the PO approximation has been developed for the problem of EM plane wave diffraction by a fully illuminated semi-infinite perfectly conducting cone [31], as illustrated in Fig. 14(a); thus, the incident angle θ measured from the axial direction must remain less than the half cone angle θ_c (see Fig. 14(a)). This plane wave diffraction solution also provides the tip diffracted ray field which propagates along the generator of the cone to any point Q on the surface. Let an electric current point source $\hat{p}_t \delta(\bar{r} - \bar{r}_p)$ at a distant point P produce this locally plane wave set of fields $(\bar{E}_t^i, \bar{H}_t^i)$ which are incident at Q_C and Q. For convenience, $\hat{p}_t$ is directed perpendicular to the ray (or local plane wave) incident at an angle θ; i.e., $\hat{p}_t = \hat{\theta}$ or $\hat{p}_t = \hat{\phi}$, where θ is shown in Fig. 14(a). Also, let $\hat{p}_t$ produce the total field $(\bar{E}_t, \bar{H}_t)$ at Q, where in the UTD sense,

$$\bar{E}_t(Q) = \bar{E}_t^i(Q) + \bar{E}_t^r(Q) + \bar{E}_t^d(Q) \tag{19}$$

$$\bar{H}_t(Q) = \bar{H}_t^i(Q) + \bar{H}_t^r(Q) + \bar{H}_t^d(Q) \tag{20}$$

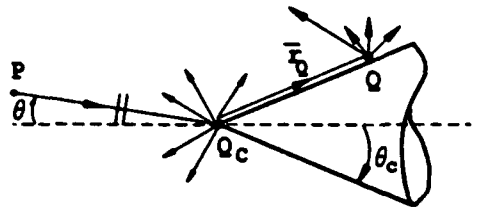
with $(\bar{E}_t^d, \bar{H}_t^d)$ representing the cone tip diffracted fields as given in [31]. The $(\bar{E}_t(Q), \bar{H}_t(Q))$ at Q also directly provides, via the reciprocity theorem, a knowledge of the fields $(\bar{E}(P); \bar{H}(P))$ radiated to the far-zone point P by a point current source $\hat{p}\delta(\bar{r} - \bar{r}_Q)$ at Q as in Fig. 14(b); thus,

$$\hat{p}_t \cdot \bar{E}(P) = \begin{cases} \hat{j} \cdot \bar{E}_t(Q), & \text{if } \hat{p} = \hat{j} \text{ where } \hat{j} \text{ is the strength of an electric current point source at } Q \\ -\hat{m} \cdot \bar{H}_t(Q), & \text{if } \hat{p} = \hat{m} \text{ where } \hat{m} \text{ is the strength of a magnetic current point source at } Q. \end{cases} \tag{21}$$

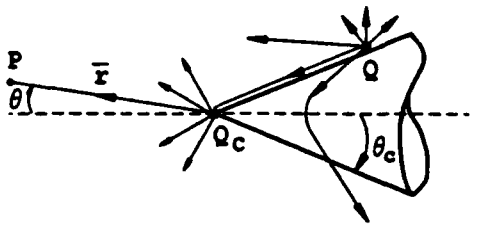
Generally, the contribution from the cone tip diffraction to the far field radiation by antennas on cones becomes negligible outside the paraxial region; this point will be clarified later when dealing with radiation from antennas on a smooth convex surface.

5) UTD Ray Fields Associated with the Diffraction by Smooth Convex Surfaces: UTD solutions for the problems of diffraction by smooth, perfectly conducting convex surfaces are useful, for example, for predicting the EM scattering from aircraft fuselage shapes or ship masts when they are illuminated by airborne or shipboard antennas, respectively, and also for predicting the EM radiation and mutual coupling associated with antennas placed conformally on smooth convex portions of an aircraft, missile or spacecraft, etc. Three separate cases are considered below.

a) Source and observation points off the smooth convex surface: The UTD solution for the case when the source (antenna) and observation points are both off the convex surface is obtained from a uniform asymptotic solution to



(a)



(b)

Fig. 14. Cone tip diffraction within the paraxial region ($\theta < \theta_c$).

the problem of EM scattering by a circular conducting cylinder [32], and it is given separately for the lit and the shadow regions. The field at P_L in the lit region is associated with the incident and reflected ray paths as shown in Fig. 15(b) and it is given by

$$\overline{E}(P_L) \sim \overline{E}^i(P_L)U + \overline{E}^{gr}(P_L)U \tag{22}$$

where $\overline{E}^i$ is the GO incident field and $\overline{E}^{gr}$ is the generalized reflected field which contains surface diffraction effects (in addition to the GO reflected field $\overline{E}^r U$). The latter effects become significant only within the transition region near the SSB shown in Fig. 3. The step function U in (22) serves as an SSB indicator:

$$U = \begin{cases} 1, & \text{in the lit region which lies above the SSB} \\ 0, & \text{in the shadow region which lies below the SSB} \end{cases} \tag{23}$$

The extent of the transition region around the SSB is of order $1/(m(Q_1))$, where

$$m(\bullet) = \left[\frac{k\rho_g(\bullet)}{2}\right]^{1/3} \tag{24}$$

and $\rho_g(\bullet)$ is the radius of curvature at any point $(\bullet)$ along the surface ray. The field $\overline{E}^{gr}$ is expressed as [32], [33]:

$$\overline{E}^{gr}(P_L) \sim \overline{E}^i(Q_R) \cdot \overline{\overline{R}} \sqrt{\frac{\rho_1^r}{\rho_1^r + s^r} \frac{\rho_2^r}{\rho_2^r + s^r}} e^{-jks^r} \tag{25}$$

with

$$\overline{\overline{R}} = \mathcal{R}_s \hat{e}_\perp \hat{e}_\perp + \mathcal{R}_h \hat{e}_\parallel^i \hat{e}_\parallel^r \quad . \tag{26}$$

The unit vectors contained in (26), and the quantities $\rho_{1,2}^r$ and s^r are the same as those given previously in (5c). The UTD functions $\mathcal{R}_s$ and $\mathcal{R}_h$ in (26) are defined in [32], [33] and they contain two transition functions, namely the F function introduced in (10) as well as the Pekeris function $\tilde{P}_{s,h}(\xi_l)$ which is well tabulated [34], in which $\xi^l = -2m(Q_R)\cos\theta$. Here θ represents the incident angle defined in Fig. 8, and $m(Q_R)$ contains $\rho_g(Q_R)$ which is the surface radius of curvature in the plane of incidence at Q_R. It is noted that $\tilde{P}_{s,h}$ contains an integral of the Airy functions [32], namely:

$$\tilde{P}_s(\delta) = \frac{e^{-j\pi/4}}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{V(\tau)}{W_2(\tau)} e^{-j\delta\tau} d\tau \tag{27a}$$

$$\tilde{P}_h(\delta) = \frac{e^{-j\pi/4}}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{V'(\tau)}{W_2'(\tau)} e^{-j\delta\tau} d\tau \tag{27b}$$

where

$$2jV(\tau) = W_1(\tau) - W_2(\tau) \tag{28a}$$

$$W_{\frac{1}{2}}(\tau) = \frac{1}{\sqrt{\pi}} \int_{\infty e^{\mp j2\pi/3}}^{\infty} \left[e^{\tau t - t^3/3}\right] dt. \tag{28b}$$

Next, the field at P_S in the shadow region is given by [32], [33]

$$\overline{E}(P_S) \sim \overline{E}^d(P_S)[1 - U] \tag{29}$$

where

$$\overline{E}^d(P_S) \sim \overline{E}^i(Q_1) \cdot \overline{\overline{T}}(Q_1, Q_2) \sqrt{\frac{\rho^d}{s^d(\rho^d + s^d)}} e^{-jks^d} \tag{30}$$

with one of the diffracted ray wavefront surface radii of curvatures, ρ^d, shown in Fig. 15; likewise, the diffracted ray distance s^d from Q_2 to P_S is also shown in that figure. The dyadic transfer coefficient $\overline{\overline{T}}$ is given as [32], [33]:

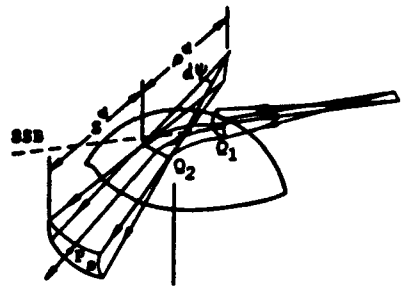
$$\overline{\overline{T}}(Q_1, Q_2) = \left[\hat{b}_1\hat{b}_2\mathcal{D}_s + \hat{n}_1\hat{n}_2\mathcal{D}_h\right] \cdot \sqrt{\frac{d\eta(Q_1)}{d\eta(Q_2)}} e^{-jkt} \tag{31}$$

in which $d\eta(\bullet)$ is the width of the surface ray tube (or strip) at any point $(\bullet)$ along the surface ray path, and t equals the arc length of the surface ray path from Q_1 to Q_2. It is noted that the surface rays constitute geodesic paths on the convex surface. It is seen from (31) that $\overline{\overline{T}}$ is expressed compactly in terms of orthogonal unit vectors $(\hat{t}, \hat{n}, \hat{b})$ fixed in the surface ray with $\hat{t}$ being a unit tangent to the surface ray and $\hat{n}$ is a unit normal to the surface along the surface ray, while $\hat{b}$ is the binormal vector $(\hat{b} = \hat{t} \times \hat{n})$. Again, $\mathcal{D}_s$ and $\mathcal{D}_h$ both contain F as in (10) as well as $\tilde{P}_{s,h}(\xi)$. The quantity ξ is sometimes referred to as the shadow Fock parameter given by

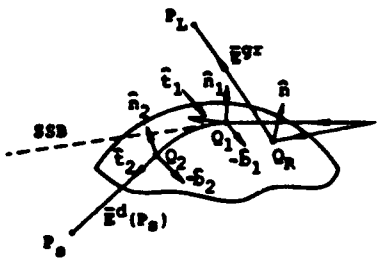
$$\xi = \int_{Q_1}^{Q_2} \frac{m(t')}{\rho_g(t')} dt'. \tag{32}$$

The parameters ξ^l, ξ, X^l and X^d present in (26) and (31), which are defined elsewhere in detail [32], [33], ensure that the total UTD field is continuous across the SSB; i.e., $\overline{E}(P_L)$ in (22) and $\overline{E}(P_S)$ in (29) are equal at the SSB.

It is noted that in the lit zone outside the SSB transition region, $\overline{E}^{gr} \to \overline{E}^r$ which is the usual GO reflected field (i.e. $\overline{\overline{R}} \to \overline{R}$ as in (5)). Likewise, in the shadow zone outside the



(a)



(b)

Fig. 15. Ray paths for scattering by a smooth convex surface.

SSB transition region, $\mathcal{D}_{s,h} \to T_{s,h}$, where $T_{s,h}$ is defined as

$$T_{s,h} \equiv \left[\sum_{n=1}^{N} D_n^{s,h}(Q_1) \cdot \exp\left(-\int_{Q_1}^{Q_2} \alpha_n^{s,h}(t')dt'\right) D_n^{s,h}(Q_2)\right] \tag{33}$$

in which $D_n^{s,h}(Q)$ is the Keller's (or GTD type) diffraction coefficient for the nth surface ray mode which indicates how the surface modes are launched at Q_1 by the incident ray which grazes the surface [35]; by reciprocity, $D_n^{s,h}$ at Q_2 indicates how the surface ray modes detach from the surface into the external medium. Similarly, $\alpha_n^{s,h}$ indicates the rate of attenuation of the surface ray modes [35] due to the continual tangential shedding of energy along the surface ray. Thus the UTD result automatically recovers the GTD result outside the SSB transition layer.

b) Radiation by sources on a smooth convex surface: Consider the radiation by a slot or a short thin wire antenna on a smooth, perfectly conducting surface. A UTD analysis of the radiation from these antennas can be constructed in terms of a uniformly asymptotic high-frequency approximation to the dyadic Green's function, $\overline{\overline{\Gamma}}_{i,m}$, which provides the radiation from a point electric (i) or magnetic (m) current source $\overline{p}\delta(\overline{r} - \overline{r}_{Q'})$ at Q' on the convex boundary.

$$\hat{p} = \begin{cases} \hat{i} & \text{for an electric point current source at } Q' \\ \hat{m} & \text{for a magnetic point current source at } Q' \end{cases} \tag{34}$$

The electric field $\overline{E}(P)$ radiated by $\hat{p}$ at Q' can then be expressed as

$$\overline{E}(P) = \begin{cases} \overline{\overline{\Gamma}}_i(P|Q') \cdot \hat{p}, & \text{if } \hat{p} = \hat{i} \\ \overline{\overline{\Gamma}}_m(P|Q') \cdot \hat{p}, & \text{if } \hat{p} = \hat{m} \end{cases} \tag{35}$$

in which $\overline{\overline{\Gamma}}_{i,m}$ is obtained from uniform asymptotic solutions to problems of radiation by $\hat{p}$ on conducting cylinders and spheres [36].

$$\overline{\overline{\Gamma}}_m \sim \begin{cases} (-jk/4\pi)\left(\hat{b}_l'\hat{n}A + \hat{t}_l'\hat{b}B + \hat{b}_l'\hat{b}C + \hat{t}_l'\hat{n}D\right)(e^{(-jks)}/s \\ \quad \text{for } P = P_L \\ (-jk/4\pi)\left(\hat{b}'\hat{n}T_1H + \hat{t}'\hat{b}T_2S + \hat{b}'\hat{b}T_3S + \hat{t}'\hat{n}T_4H\right) \\ \quad \cdot e^{-jkt}\sqrt{(d\psi_o/d\eta(Q))}[\rho_g(Q)/\rho_g(Q')]^{1/6} \\ \quad \cdot\sqrt{\rho^d/(s^d(\rho^d + s^d))}\exp(-jks^d), \\ \quad \text{for } P = P_S \end{cases} \tag{36}$$

and

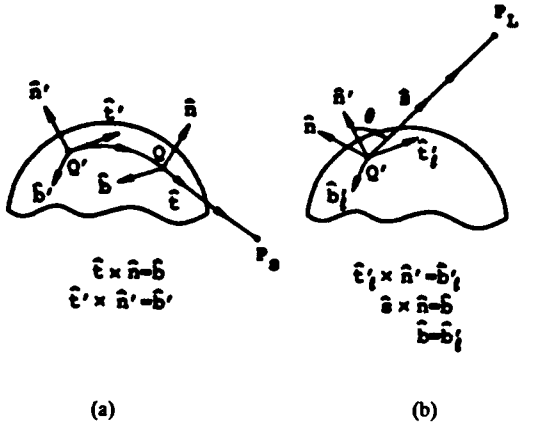
$$\overline{\overline{\Gamma}}_i \sim \begin{cases} (-jkZ_o/4\pi)\left(\hat{n}'\hat{n}M + \hat{n}'\hat{b}N\right)(e^{-jks}/s), \\ \quad \text{for } P = P_L \\ (-jkZ_o/4\pi)\left(\hat{n}'\hat{n}T_5H + \hat{n}'\hat{b}T_6S\right) \\ \quad \cdot e^{-jkt}\sqrt{(d\psi_o/d\eta(Q))}[(\rho_g(Q)/\rho_g(Q'))]^{1/6} \\ \quad \cdot\sqrt{\rho^d/(s^d(\rho^d + s^d))}\exp(-jks^d), \\ \quad \text{for } P = P_S. \end{cases} \tag{37}$$

The field point $P = P_L$ in the lit region (where the source at Q' is directly visible), and $P = P_S$ in the shadow region. Although the fields in (36) and (37) are given separately for $P = P_L$ and $P = P_S$, respectively, they join smoothly at the shadow boundary SSB which is defined by a plane tangent to the surface at Q'. The quantities ρ^d and s^d are shown in Fig. 15; also, $d\eta(Q)$ is the width of the surface ray strip at Q. The $d\eta$ was defined earlier in (31); furthermore, $d\psi_o$ and $d\psi$ are the angles subtended by the surface ray strip at Q' and at Q, respectively (e.g., $d\psi$ is shown in Fig. 15) [36]. The unit vectors $(\hat{t}', \hat{b}', \hat{n}')$ at Q' and $(\hat{t}, \hat{n}, \hat{b})$ at Q are fixed in the surface ray from Q' to Q as in Fig. 16, and they have the same meaning as in Fig. 15. Similarly, the unit vectors $(\hat{t}_l', \hat{n}, \hat{b}_l')$ are fixed in the ray from Q' to P_L such that $\hat{t}_l'$, $\hat{n}$, $\hat{t}'$ and $\hat{n}'$ all lie in the plane of incidence (defined by $\hat{s} = \overline{(Q'P_L)}/(|\overline{Q'P_L}|)$ and $\hat{n}'$) and $\hat{n} \cdot \hat{s} = 0 = \hat{b}_l' \cdot \hat{s}$ as shown in Fig. 16. The usual angle of incidence $\theta = \cos^{-1}(\hat{n}' \cdot \hat{s})$ defines the radiation angle measured from the $\hat{n}'$ direction.

The quantities A, B, C, D, M and N in (36) and (37) for $P = P_L$, and the quantities T_1 through T_6 in those equations for $P = P_S$ are all defined in [36]; they contain the special UTD transition functions $g(\bullet)$ and $\tilde{g}(\bullet)$ corresponding to the well tabulated radiation Fock functions [34], [36] that are expressed in terms of an integral of Airy functions.

$$g(\delta) = \frac{1}{\sqrt{\pi}} \int_{\infty e^{-j(2\pi/3)}}^{\infty} d\tau e^{-j\delta\tau}[W_2'(\tau)]^{-1} \tag{38a}$$

$$\tilde{g}(\delta) = \frac{1}{\sqrt{\pi}} \int_{\infty e^{-j(2\pi/3)}}^{\infty} d\tau e^{-j\delta\tau}[W_2(\tau)]^{-1}. \tag{38b}$$



(a) (b)

Fig. 16. Unit vectors fixed in rays to P_L and P_S from a source on a convex surface.

The argument δ of the Fock functions is given by $\xi_l = -m(Q')\cos\theta$ for $P = P_L$, while it is given by the shadow Fock parameter $\xi = \int_{Q'}^{Q} (m(t')/\rho_g(t'))dt'$ when $P = P_S$ as defined previously in (32).

Outside the SSB transition region, where $\xi_l \ll 0$ and $\xi \gg 0$ the UTD results in (36) and (37) automatically reduce to the GTD form; namely, $A \to 2$, $B \to 2\cos\theta$, $M \to 2\sin\theta$ and $(C, D, N) \to 0$ in (36) and (37) for $P = P_L$ and $\xi_l \ll 0$, and similarly, the results for $P = P_S$ and for $\xi \gg 0$ reduce to terms involving

$$\sum_{n=1}^{N} L_n^{s,h}(Q')\exp\left(-\int_{Q'}^{Q} \alpha_n^{s,h}(t')dt'\right) D_n^{s,h}(Q).$$

In the latter sum, the $L_n^{s,h}(Q')$ (which is proportional to $D_n^{s,h}(Q')$) is the launching coefficient at Q' of the nth surface ray mode, and $D_n^{s,h}(Q)$ is the nth surface ray mode diffraction coefficient introduced earlier in (33). Within the SSB transition region, the GTD launching and diffraction mechanisms are no longer distinct; indeed, such a coupling between the launching and diffraction effects within the transition region is naturally contained in the UTD results of (36) and (37) due to the presence of $g(\delta)$ and $\tilde{g}(\delta)$ in those equations.

The geodesic surface ray path from Q' to Q may be torsional. A torsional path, is one for which $\hat{b} \neq \hat{b}'$; i.e., a torsional path is a nonplanar curve. It is noted that the geodesic surface ray paths are helices on convex cylinders, and they are great circles on spheres; they can be found easily for developable surfaces, but they must be found numerically for more general surfaces such as spheroids, etc. [36]. Furthermore, for closed surfaces, rays that creep around or encircle such surfaces can also diffract (shed) tangentially into the lit region. Generally, for electrically large closed surfaces the contribution from such encircling rays is weak because of the continuous tangential shedding of energy along such rays.

The result in (35) can be readily generalized to deal with the radiation from a slot antenna on a convex surface [36]; thus, the field radiated by a slot antenna becomes (via a direct generalization of (21) with $\hat{p} = \hat{m}$):

$$\overline{E}(P) = \int\int_{s_a} \overline{\overline{\Gamma}}_m(P|Q') \cdot [\overline{M}_S(Q')]ds' \qquad (39)$$

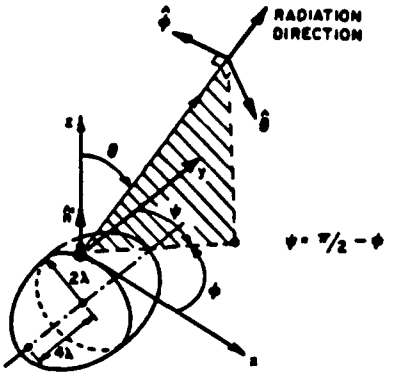
where $\overline{M}_S(Q') = \overline{E}_a(Q') \times \hat{n}'$ is the equivalent magnetic current in terms of the transmitting electric field $\overline{E}_a(Q')$ in the slot aperture of area S_a; this $\overline{M}_S$ replaces the aperture S_a which is now short circuited. Likewise, the radiation from a short thin monopole of height h and transmitting current $I(l')$ fed at the base Q' on a convex surface can be found as [36]:

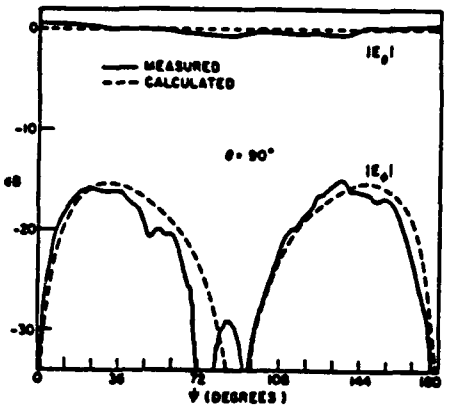
$$\overline{E}(P) \approx \begin{cases} \overline{\overline{\Gamma}}_i(P|Q') \cdot \hat{n}' \int_0^h I(l') \\ \qquad \cdot \cos(kl'\cos\theta)dl', & \text{if } P = P_L \\ \overline{\overline{\Gamma}}_i(P|Q') \cdot \hat{n}' \int_0^h I(l')dl', & \text{if } P = P_S. \end{cases} \qquad (40)$$

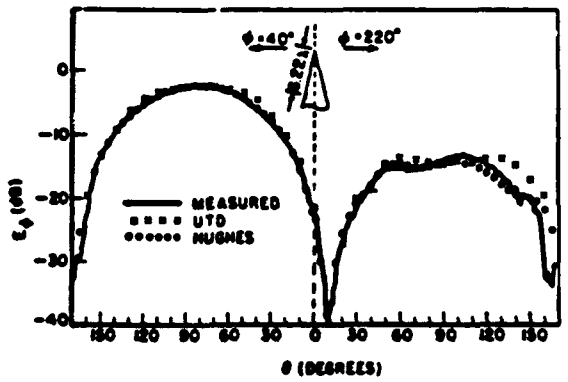
Figure 17(a) indicates the far zone radiation pattern of a short, thin monopole antenna on a spheroid, which is calculated in the SSB plane (i.e. in the plane tangent to the spheroid at the base of the monopole) via UTD, and is shown to compare very well with measurements. Besides the constant $|\overline{E}_\theta|$ pattern which is the only component that would exist in the SSB plane for a monopole on a sphere, or on a finite or infinite flat ground plane, there is an additional (cross-polarized) $|\overline{E}_\phi|$ component of the pattern that is present for the spheroid because it has two different principal surface curvatures K_1 and K_2; i.e., $\overline{E}_\phi$ in the SSB plane of the spheroid results from the "launching" of torsional surface rays by the monopole since torsion is proportional to $|K_1 - K_2|$. The present UTD solution thus predicts the complex, surface dependent field and polarization effects in the SSB transition region through the explicit presence of torsion factors and the radiation Fock functions in (36) and (37). It is noted that, even though Fock functions are utilized here, Fock's original work did not contain effects of torsional surface rays. The far zone UTD radiation pattern of a radial slot on a semi-infinite cone is seen to compare very well with an exact modal (eigenfunction) solution in Fig. 17(b). The effect of the tip is ignored in Fig. 17(b); however, as pointed out earlier, this effect is generally negligible outside the paraxial region.

c) Mutual coupling between antennas on a convex surface: The UTD expressions for the EM fields $(\overline{E}(Q), \overline{H}(Q))$ at Q on a convex surface that are produced by a slot antenna, or a short and thin monopole antenna on the same surface, respectively, are given in detail in [37]; those analytical expressions have been obtained from the high-frequency solutions to the same canonical problems as for the radiation problem in part (b) above, and are presented only symbolically in (41) and (42), shown on the next page.Expressions alternative to those in [34] for the slot case are presented in [38].

The UTD expressions for $\overline{\overline{\Gamma}}_{ee,eh,he,hh}$ which occur in (41) and (42) contain special transition functions $\tilde{U}(\xi)$ and $\tilde{V}(\xi)$ [37], [38]; these transition functions are expressed in terms of an integral containing a ratio of Airy functions, and they keep the above solutions valid in their SSB transition







$|E_\phi|$ radiation pattern of a radial slot in a cone

Fig. 17. Radiation patterns of antennas on perfectly conducting spheroids and cones (cone half angle = 10°).

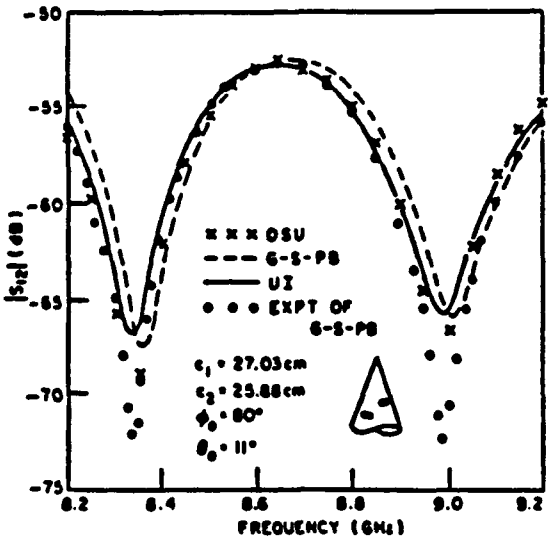


Fig. 18. Mutual coupling between a pair of slots on a perfectly conducting cone.

region on the surface.

$$\tilde{U}(\xi) = \left[\frac{kt}{2m(Q')m(Q)\xi}\right]^{3/2} \cdot \frac{\xi^{3/2}e^{j(3\pi/4)}}{2\sqrt{\pi}} \cdot \int_{\infty e^{-j(2\pi/3)}}^{\infty} d\tau e^{-j\xi\tau} \frac{W_2'(\tau)}{W_2(\tau)} \tag{43}$$

$$\tilde{V}(\xi) = \left[\frac{kt}{2m(Q')m(Q)\xi}\right]^{1/2} \cdot \frac{\xi^{1/2}e^{j\pi/4}}{2\sqrt{\pi}} \int_{\infty e^{-j(2\pi/3)}}^{\infty} d\tau e^{-j\xi\tau} \frac{W_2(\tau)}{W_2'(\tau)}. \tag{44}$$

Expressions for the mutual coupling between a pair of antennas on a convex surface can be found using (43) and (44) as described in [37], [38]. Figure 18 indicates the mutual coupling between a pair of slot antennas on a cone calculated via UTD [37]; it is seen to compare very well with an exact eigenfunction solution. The pattern in Fig. 18 results from the interference between the dominant surface rays and the tip diffracted ray. The tip diffracted ray field is calculated as described in [38].

The UTD edge and convex surface diffraction solutions discussed above are employed to predict the radiation patterns of a TACAN monopole antenna mounted behind the canopy on the top side of an F-16 aircraft fuselage; the

$$\left\{ \begin{matrix} \bar{E}(Q) \\ \bar{H}(Q) \end{matrix} \right\} \sim \left\{ \begin{matrix} \bar{\bar{\Gamma}}_{ee}(Q|Q') \\ \bar{\bar{\Gamma}}_{he}(Q|Q') \end{matrix} \right\} \cdot \hat{n}' \int_o^h I(l')dl', \quad \begin{matrix} \text{for a monopole} \\ \text{antenna as in (40)} \end{matrix} \tag{41}$$

and

$$\left\{ \begin{matrix} \bar{E}(Q) \\ \bar{H}(Q) \end{matrix} \right\} \sim \int\int_{Sa} \left\{ \begin{matrix} \bar{\bar{\Gamma}}_{eh}(Q|Q') \\ \bar{\bar{\Gamma}}_{hh}(Q|Q') \end{matrix} \right\} \cdot [\bar{E}_a(Q') \times \hat{n}']ds', \quad \begin{matrix} \text{for a slot antenna} \\ \text{as in (39)} \end{matrix} \tag{42}$$

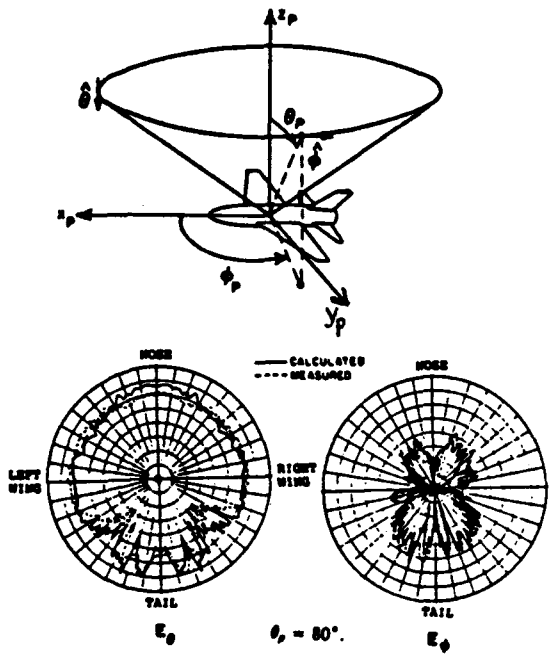


Fig. 19. Radiation pattern of monopole antenna on an F-16 fighter aircraft (see [39]).

results are shown in Fig. 19 [39]. The UTD model of the F-16 is built up from connecting spheroidal and flat plate surfaces. The excellent comparison of these UTD results with measurements is also shown in that figure.

An application of the UTD solution for radiation and mutual coupling associated with antennas on a convex surface is shown in Fig. 20 for predicting the radiation from a 9 × 9 element dominant mode rectangular waveguide-fed axial slot antenna phased array in a perfectly conducting cylinder; this array exhibits a cosine tapered distribution along both the axial and circumferential directions. The cosine taper is realized incorporating the effects of mutual coupling. Figure 20 shows the radiation pattern of this array when it is phased to radiate in the $\theta = 45°$, $\phi = 45°$ direction [40].

6) Uniform Analysis of Reflection within Caustic Regions: Figure 6 illustrates a concave–convex surface of revolution which contains an inflection point along its generator; such a surface can occur in the design of shaped subreflectors in dual reflector antenna systems. The feed, which is a source of a spherical wave, illuminates the subreflector which is assumed to be in the far zone of the feed. The rays reflected from the subreflector form a smooth caustic surface of revolution. It is usually of interest to find the fields scattered by the subreflector which then illuminate the main reflector.

One can employ GO to find the rays reflected from the subreflector as shown in Fig. 6. On the lit side of the caustic there are two real GO reflected rays that contribute to the field at P_L. However, conventional GO fails to predict a field at P_S on the shadow side of the caustic where no real

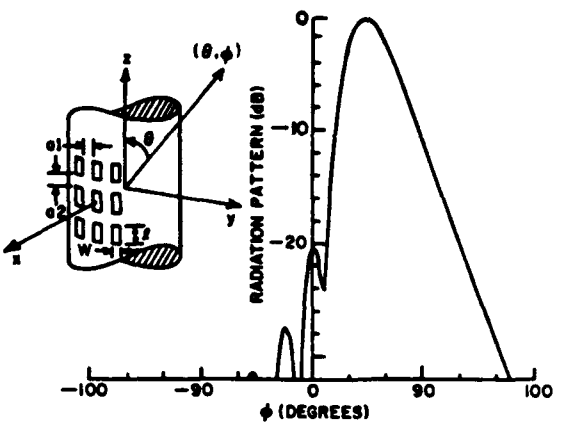


Fig. 20. Radiation from a 9 × 9 axial slot phased array on a perfectly conducting conducting circular cylinder. $a = 5\lambda$; $a1 = 0.114\lambda$; $a2 = 0.095\lambda$. Slot length = 0.686λ; slot width = 0.305λ.

reflected rays exist, and it also fails at the caustic where it predicts a field singularity. Therefore, conventional GO must be patched up by a uniform GO solution which not only provides a bounded and smooth variation of the field across the caustic, but which also automatically recovers the real ray fields of GO on the lit region outside the caustic transition layer, and which likewise recovers the "complex" ray field [41] outside the transition layer on the shadow side of the caustic (since "real" ray fields cannot exist there). Such a uniform GO solution for the scattered fields $(\overline{E}^s, \overline{H}^s)$ is described in [14]; it is summarized below:

$$\overline{E}^s(P_L) \sim 2\pi j e^{-jk\delta_l}\left[\overline{P}_l k^{-1/3} Ai(-k^{2/3}\zeta_l) + j^{\frac{1}{2}}\zeta_l \overline{Q}_l k^{-2/3} Ai'(-k^{\frac{2}{3}}\zeta_l)\right] \quad (45)$$

where Ai is an Airy function [13]–[15] given by

$$Ai(\delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, \exp(-j(t^3/3 + \delta t))$$

and Ai' is its derivative. Furthermore,

$$\begin{Bmatrix}\overline{P}_l \\ \overline{Q}_l\end{Bmatrix} = \frac{\zeta_l^{1/4}}{2}\sqrt{\frac{k}{\pi}}\left[\frac{\overline{A}(Q_a)\cdot\overline{\overline{R}}(Q_a)}{s_a^i}\cdot\sqrt{\frac{\rho_{1a}^r}{\rho_{1a}^r + s_a^r}}\sqrt{\frac{\rho_{2a}^r}{\rho_{2a}^r + s_a^r}}e^{-j\pi/4}\right.$$
$$\left. \mp \frac{\overline{A}(Q_b)\cdot\overline{\overline{R}}(Q_b)}{s_b^i}\sqrt{\frac{\rho_{1b}^r}{\rho_{1b}^r + s_b^r}}\sqrt{\frac{\rho_{2b}^r}{\rho_{2b}^r + s_b^r}}e^{+j\pi/4}\right] \quad (46a)$$

in which

$$\delta_l = \frac{1}{2}[(s_b^i + s_b^r) + (s_a^i + s_a^r)] \quad (46b)$$

$$\zeta_l^{\frac{3}{2}} = \frac{3}{4}[(s_b^i + s_b^r) + (s_a^i + s_a^r)] \quad (46c)$$

and

$$\overline{E}^i(Q_{a,b}) = \overline{A}(Q_{a,b})\frac{\exp(-jks^i_{a,b})}{s^i_{a,b}} \quad \begin{pmatrix} \text{field incident at} \\ Q_{a,b} \text{ from the feed} \end{pmatrix}. \tag{46d}$$

It is noted that $\overline{\overline{R}}$ in (46a) is the dyadic reflection coefficient as in (5). While (45) is given for the lit side, a similar expression exists for the shadow side, namely,

$$\overline{E}^s(P_s) \sim 2\pi j e^{-jk\delta_s}\Big[\overline{P}_s k^{-1/3} Ai(+k^{2/3}\zeta_s) + \zeta_s^{-1/2}\overline{Q}_s k^{-2/3} Ai'(+k^{\frac{2}{3}}\zeta_s)\Big]. \tag{47}$$

In (46d), the Q_a and Q_b correspond to the two "real" points of reflection on the surface as in Fig. 6; the parameters in (46b) and (46c) are defined in terms of Q_a and Q_b. The parameters in (47), for the shadow side, are defined in terms of "complex" points of reflection Q_{ac} and Q_{bc}; these complex points are determined by an analytical continuation of the original surface into complex coordinate space [14], [41]. In practice, the subreflector in Fig. 6 is bounded by an edge which then truncates the caustic surface in question. The results in (45) and (47) are valid for observation points near the smooth portion of the caustic away from the caustic truncation and away from the second (or the other) caustic surface which also exists. Outside the given caustic transition region, (45) automatically recovers the GO result.

7) Multiple Ray Interactions: A diffracted ray which is incident on a discontinuity undergoes a second diffraction to create a doubly diffracted ray. Likewise, doubly diffracted rays can produce triply and higher order multiply diffracted rays. The effect of multiply diffracted rays is generally quite weak and may be ignored in that case. However, one can easily assess the importance of the latter, because leaving these out generally creates a discontinuity in the field (much like GO exhibits discontinuities along GO ray shadow boundaries); if this discontinuity is significant then it is clear that the multiple interactions must be included to some order until the discontinuity becomes sufficiently small. Finally, rays reflected and then diffracted (or vice versa) are of the same order as singly diffracted rays; thus they must be generally included to keep all significant interactions to the same order of asymptotic approximation (in terms of inverse powers of k) [3]. Multiple interactions within ray transition regions need to be treated with care [63].

B. ECM

An expression for the GTD/UTD edge diffracted field has been presented in (8c) above, namely,

$$\overline{E}^d(P) \sim \overline{E}^i(Q_E)\cdot\overline{\overline{D}}_e\sqrt{\frac{\rho^d}{s^d(\rho^d+s^d)}}e^{-jks^d}.$$

This expression reveals that the edge diffracted field has a singularity at the edge where $s^d = 0$; such a singularity results from the fact that the edge is a caustic of the edge diffraction rays, and this caustic at Q_E is evident from the edge diffracted ray tube illustrated in Fig. 9. The GTD/UTD expression for the asymptotic high-frequency ray field is valid away from the edge (i.e., it is valid outside the so-called edge boundary layer), and the proper behavior near the edge must be obtained from separate considerations. On the other hand, if $\rho^d < 0$ then the other diffracted ray caustic can occur in the external space surrounding the wedge whenever the observation point at P is such that $s^d = |\rho^d|$, and the expression for $\overline{E}^d(P)$ of (8c) thus becomes singular and consequently fails at and near this caustic; such a caustic can generally occur along the diffracted ray if the edge is curved or if the incident wavefront is concave. For smooth caustics of diffracted rays one could use the expressions in (45) and (47) directly within the diffracted ray caustic region except that the reflected ray parameters present in (45) and (47) must now be replaced by the corresponding diffracted ray parameters. However, the use of the ECM in this case will yield the same result as in (45) and (47) if the integrals present in the ECM, which are defined later on, are evaluated using a uniform asymptotic procedure. Furthermore, the ECM is very useful for treating a point caustic of diffracted rays (as in Fig. 5 for a symmetric parabolic reflector with the feed at the focus); the uniform approximation of (45) and (47) is not valid in regions at and near the intersection or proximity of the two smooth caustic surfaces, nor where these two caustic surfaces degenerate to form a single point caustic.

The basic idea behind ECM may be understood as follows. If $ks^d \gg 1$ but $s^d \ll |\rho^d|$, then in the near zone of the edge but sufficiently far from Q_E, the expression for $\overline{E}^d(P)$ in (8c) becomes:

$$\overline{E}^d(P) \sim \overline{E}^i(Q_E)\cdot\overline{\overline{D}}_e\frac{e^{-jks^d}}{\sqrt{s^d}}. \qquad \begin{pmatrix} \text{with} \\ s^d \ll |\rho^d| \\ ks^d \gg 1 \end{pmatrix} \tag{48}$$

Clearly the field at P in (48) may be viewed as being produced by an appropriate equivalent line source tangent to the curved edge at Q_E, because a line source field also exhibits an symptotic behavior of the type $e^{-jks^d}(s^d)^{-1/2}$ as in (48) when $ks^d \gg 1$, to describe a cylindrical wave as illustrated in Fig. 5(b). Thus one can find the strengths of equivalent electric (I) and magnetic (M) line currents locally tangent to the edge (i.e., along $\hat{e}$) at Q_E, which generate the desired fields $(\overline{E}^d(P), \overline{H}^d(P))$. For a perfectly conducting edge, the equivalent line currents I and M are given by [16]

$$\begin{Bmatrix} I(Q_E) \\ M(Q_E) \end{Bmatrix} = \frac{-e^{-j\pi/4}}{\sin\beta_0}\sqrt{\frac{8\pi}{k}}\begin{Bmatrix} Y_0 \\ Z_0 \end{Bmatrix}\hat{e} \cdot \begin{Bmatrix} \overline{E}^i(Q_E)D_{es} \\ \overline{H}^i(Q_E)D_{eh} \end{Bmatrix} \tag{49}$$

in which $D_{es,eh}$ have been indicated previously in (9) and are evaluated in (49) for a diffracted ray which lies on the Keller cone and in the caustic direction. Only if the phase of $\overline{E}^i(Q_E)$ in (48) is uniform then does (48) describe a "locally" cylindrical wave emanating from the

edge as in Fig. 5(b). On the other hand, if the incident rays strike the edge obliquely (so $\beta_o \neq \pi/2$) then the phase of $\overline{E}^i(Q_E)$ is not uniform, and neither does the phase of I and M remain constant but instead contains a traveling wave factor automatically through the presence of $\overline{E}^i(Q_E)$ in (49). In the latter case the diffracted field behavior in (48), and likewise the corresponding asymptotic line source field behavior, $e^{jks^d}(s^d)^{-1/2}$, now describes a more general conical rather than a cylindrical wave. In the ECM, these conical waves thus locally simulate the Keller cones of edge diffracted rays. Even though (8c) becomes singular at diffracted ray caustics, the currents in (49) are defined and well behaved at every point along the curved edge, and hence they can be incorporated within the radiation integral to yield a bounded result for the total diffracted field $\overline{E}^d_{\text{total}}$ at and near the caustic. Thus

$$\overline{E}^d_{\text{total}}(P) \sim \frac{jkZ_o}{4\pi} \oint \left[\hat{R} \times \hat{R} \times I\hat{e} + Y_o \hat{R} \times M\hat{e}\right] \cdot \frac{e^{-jKR}}{R} dl' \tag{50}$$

where $\overline{R}$ is the vector from Q_E to P, and the integration is around the edge contour which produces the caustic of diffracted rays.

It is noted that an edge diffracted ray exhibits the local line source field variation of the type $e^{jks^d}(s^d)^{-1/2}$ in (48) only when $\overline{\overline{D}}_e$ is not range dependent; i.e., only when one observes the edge diffracted field outside the edge boundary layer and external to the incident and reflection boundary (ISB and RSB) transition regions where the UTD reduces to GTD. This is true because the special range dependent Fresnel type UTD transition function F in $\overline{\overline{D}}_e$, which is different from unity within the ISB and RSB transition regions, modifies the $e^{jks^d}(s^d)^{-1/2}$ type cylindrical or conical wave behavior within these transition layers. Consequently, the GTD-based ECM remains valid only if the edge diffracted ray caustic transition layer does not overlap with the ISB and RSB transition layers.

The ECM is an outgrowth of some early work in [42] which was later formulated in terms of the GTD in [16] to yield (49). A heuristic modification to extend the use of $D_{es,eh}$ in (49), which are defined only on the Keller cone, so that they can be approximately generalized to be defined along radiation directions lying outside the Keller cone is provided in [17] by splitting the $(\sin\beta_o)^{-1}$ factor in (49) as well as the one present in the $D_{es,eh}$ of (49), symmetrically into $\sqrt{\sin\beta'_o \sin\beta_o}$, where β'_o is the angle between the incident ray and $\hat{e}$ at Q_E, and β_o is the angle between the observation direction and $\hat{e}$ at Q_E. If $\beta_o = \beta'_o$, only then does the direction of radiation from I or M at Q_E coincide with the diffracted ray from Q_E that lies on the Keller cone. Such a generalization involving a symmetric split is useful in that away from the caustic transition layer, where the GTD is valid, it allows the integral in (50) to reduce asymptotically (i.e., for large radius of curvature of the edge) to the expected GTD description [4,17] in terms of a superposition of isolated edge diffracted ray contributions $\sum_{j=1}^{J} \overline{E}^d_j(P)$, where each term $\overline{E}^d_j(P)$ is of the type in (8c). The effect of truncating the limits of integration to the portion of the edge which is directly illuminated may create spurious contributions; this aspect and possible remedies are discussed in [4]. It is noted that unlike tr[illegible]rents, the I and M in (49), together with the [illegible]n of $\sin\beta_o$ to $\sqrt{\sin\beta'_o \sin\beta'_o}$, depend on the r[illegible] or observation direction.

The GTD-based ECM discussed above provides the diffracted field contribution without ha[illegible]g to find the diffracted ray paths as in the G[illegible] ,wever, the ECM requires an integration, which [illegible] some special cases can be evaluated in closed form, [illegible] which in general must be evaluated numerically. This ECM can be used to find the fields diffracted within the rear axial caustic region of the symmetric parabolic reflector of Fig. 5(a) as shown in Fig. 11. An analogous ECM application is to calculate the fields in the rear axial caustic direction of a coaxial waveguide fed aperture in a finite circular ground plane [43]. An ECM analysis of the radiation by an axial monopole on a circular ground plane, and on a flat-backed cone may be found in [9], and [44], respectively.

The GTD-based ECM can also be employed to describe the fields diffracted by an offset fed parabolic reflector for those observation directions in which isolated points of edge diffraction which move on the elliptic rim can coalesce and thereby create a singularity in the conventional GTD calculation. In this case, the ECM integral could asymptotically be expressed in terms of a parabolic cylinder function, thus providing an analytical result if desired.

In addition, the ECM can be employed in special cases to evaluate the fields at caustics of reflected rays, and of surface diffracted rays. Furthermore, it can be extended to treat the scattering by a class of interior waveguide discontinuities. The equivalent currents for interior waveguide regions are defined via the concept of modal ray fields which are found either exactly or asymptotically from the interior waveguide modes [45], [46]. The equivalent currents I and M, which "replace" the interior structure, and asymptotically produce the same interior modal fields as those created by the discontinuity via not only I and M, but also their images due to the effect of the waveguide walls, have been developed in [46] to find the modal reflection coefficients and the radiation by an incident modal field associated with an open-ended semi-infinite parallel plate waveguide antenna geometry. Other related work may be found in [47]–[49]. An alternative approach in which only equivalent magnetic currents M are impressed at the discontinuity and which radiate the desired fields within interior (or exterior) regions in the "presence" of the interior waveguide geometry have been developed in [50], [51]; such an ECM-based approach, which does not require one to explicitly find the images of the equivalent sources and their fields, can treat a somewhat more general class of waveguide discontinuities. Examples of the use of the latter ECM for interior regions are illustrated for finding the modal reflection coefficients of a waveguide fed horn antenna in [50], and for an open-ended circular waveguide

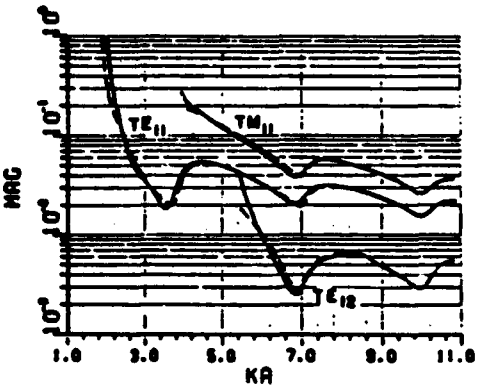


Fig. 21. Application of ECM to find the modal reflection coefficient of open-ended circular waveguide and horn antennas. (Exact Wiener–Hopf solution in: Weinsten, *The Theory of Diffraction and the Factorization Method*, Golem Press, 1968.)

aperture antenna in [51] as presented in Fig. 21. The ECM based results in Fig. 21 are seen to compare extremely well with exact Wiener–Hopf calculations.

More recently, the GTD-based ECM for edged bodies has been formulated in [52] directly from the asymptotic treatment of the integral representation for the canonical wedge diffraction problem, from which a set of slightly improved equivalent currents I and M can be identified. It may be remarked that the ECM concept is closely related to the incremental length diffraction coefficient (ILDC) concept developed by Mitzner [53]; a comparison of ILDC and ECM is available in [54], [55].

C. PTD

As indicated in Section I, the PTD was developed by Ufimtsev [19] at about the same time Keller developed the GTD. The PTD serves to correct PO, while GTD provides a correction to GO. Thus the PTD field is a superposition of the PO field and its correction which is the so-called "edge wave field." The PO field is produced by the GO approximation for the currents induced on the radiating object, whereas the edge wave field is produced by the diffracted component of the current on the radiating object. Since GTD/UTD is the sum of the GO and diffracted ray fields, it is not surprising that if the PTD radiation integrals (i.e., the PO integral plus the integral over the diffracted current component) are evaluated using high-frequency asymptotics then the PTD reduces to the GTD. Furthermore, when the asymptotics is performed in a uniform fashion, the PTD can recover the UTD. Clearly, therefore, the PTD can be employed to patch up GTD/UTD in regions where GTD/UTD and even the GTD-based ECM fails. Elsewhere, the GTD/UTD and the GTD–ECM become applicable and are expected to be far more efficient than the PTD which generally requires the evaluation of PO integrals over an electrically large radiating object. Furthermore, multiple wave interactions can generally be accounted for in a straightforward fashion using the GTD ray technique, which is not true for the PTD. Also, the PTD has been developed only for an edge at the present time; consequently, the GO current discontinuity at the geometric shadow boundary on the smooth portion of a scatterer can induce a spurious diffraction contribution to the PO integral. Ways to remove such spurious effects are discussed in [56]. Nevertheless, the PTD becomes particularly useful for patching up the GTD/UTD in regions where there is a confluence of reflected and/or diffracted ray caustic transition regions with the GO (incident or reflected) ray shadow boundary transitions regions. It is recalled that the GTD–ECM-based approach also fails there. The PTD electric field at an observation point P can be written as

$$\overline{E}(P) \sim \overline{E}^i + \overline{E}^S_{PTD} \tag{51a}$$

$$\overline{E}^S_{PTD} = \overline{E}^S_{PO} + \overline{E}^S_U \tag{51b}$$

where $\overline{E}^i$ is the classical incident field from the primary source radiating in the absence of any scattering structure, and $\overline{E}^S_{PTD}$ is the PTD based asymptotic approximation to the field scattered by the presence of the structure when excited by the primary source. Unlike the incident GO field $\overline{E}^i U_i$ which is discontinuous (see (4a)), the $\overline{E}^i$ in (51a) is continuous everywhere. The $\overline{E}^S_{PTD}$ is calculated by superposing the physical optics contribution, $\overline{E}^S_{PO}$ and the Ufimtsev correction, $\overline{E}^S_U$ as in (51b), where

$$\overline{E}^S_{PO}(P) = \frac{-jkZ_o}{4\pi} \int\int_{S_{\text{lit}}} \overline{J}^{GO}_S(\bar{r}') \cdot \left[\bar{\bar{I}} + \frac{\nabla\nabla}{k^2}\right] \frac{e^{-jk|\bar{r}-\bar{r}'|}}{|\bar{r}-\bar{r}'|} ds' \tag{52}$$

in which $\overline{J}^{GO}_S(\bar{r}')$ is the GO approximation to the current induced at any point $\bar{r}'$ on a perfectly conducting boundary excited by a primary source (the source of $\overline{E}^i$). The boundary may be a host structure for an antenna which serves as a primary source, or, the primary source could be a feed antenna for a reflecting boundary (e.g., a parabolic reflector). Thus $\overline{J}^{GO}_S(\bar{r}') = \hat{n}' \times [\overline{H}^i(\bar{r}')U_i + \overline{H}^r(\bar{r}')U_r]$ on the part of the boundary surface S_{lit} which is directly illuminated by the primary source, and $\overline{J}^{GO}_S = 0$ elsewhere (i.e., in the shadow region) on the boundary. Here, $\overline{H}^i(\bar{r}')U_i$ and $\overline{H}^r(\bar{r}')U_r$ are the incident and reflected GO components of the magnetic field at $\bar{r}'$. The position vector to the observation point is $\bar{r}$, and $\hat{n}'$ is the unit outward normal vector to the boundary at $\bar{r}'$. It appears that Ufimtsev found $\overline{E}^S_U$ indirectly in his original work [19]. If the PTD integrals in (51), (52) are approximated asymptotically, then

$$\overline{E}^S_{PO} \sim -\overline{E}^i(1-U_i) + \overline{E}^r U_r + \overline{E}^i(Q_E) \cdot \overline{D}^{PO}_e \sqrt{\frac{\rho^d}{s^d(\rho^d+s^d)}} e^{-jks^d} \tag{53}$$

and $\overline{E}^S_U(P)$ can be expressed as

$$\overline{E}^S_U(P) \sim \overline{E}^i(Q_E) \cdot \overline{D}^U_e \sqrt{\frac{\rho^d}{s^d(\rho^d+s^d)}} e^{-jks^d} \tag{54}$$

where $\overline{\overline{D}}_e^{PO}$ is identified as a PO based edge diffraction coefficient, and $\overline{\overline{D}}_e^{U}$ is a Ufimtsev edge diffraction coefficient. Actually, there can be several edge diffraction contributions to (51) and (52); however, only a single such contribution is indicated in (53) for convenience. It is interesting to observe that [17], [57]

$$\overline{\overline{D}}_e^{PO} + \overline{\overline{D}}_e^{U} = \overline{\overline{D}}_e \tag{55}$$

as one might expect. Equation (55) essentially illustrates the connection between PTD and GTD/UTD. Following (9), one can also express $\overline{\overline{D}}_e^{PO}$ and $\overline{\overline{D}}_e^{U}$ as

$$\begin{aligned}\overline{\overline{D}}_e^{PO} &= -\hat{\beta}_o'\hat{\beta}D_{es}^{PO}(\phi,\phi';\beta_o) \\ &\quad - \hat{\phi}'\hat{\phi}D_{eh}^{PO}(\phi,\phi';\beta_o) \qquad (56a)\\ \overline{\overline{D}}_e^{U} &= -\hat{\beta}_o'\hat{\beta}_o D_{es}^{U}(\phi,\phi';\beta_o) \\ &\quad - \hat{\phi}'\hat{\phi}D_{eh}^{U}(\phi,\phi';\beta_o). \qquad (56b)\end{aligned}$$

The $\overline{\overline{D}}_e^{U}$ essentially describes Ufimtsev's edge (fringe) wave diffraction pattern. While Ufimtsev found $\overline{E}_U^S$ via indirect considerations in [19], one could in retrospect employ an approximate procedure following the GTD-based ECM ideas contained in (49) and (50), as well as in the discussion below (50), to obtain $\overline{E}_U^S$; thus

$$\overline{E}_U^S(P) \approx \frac{jkZ_o}{4\pi}\oint\left[\hat{R}\times\hat{R}\times I^U\hat{e} + Y_o\hat{R}\times M^U\hat{e}\right]\cdot\frac{e^{-jkR}}{R}dl' \tag{57}$$

where the Ufimtsev type equivalent currents I^U and M^U in (57) are given by

$$\begin{aligned}&\begin{Bmatrix} I^U(Q_E) \\ M^U(Q_E)\end{Bmatrix} \\ &\quad = -\frac{e^{-j\pi/4}}{\sqrt{\sin\beta_o\sin\beta_o'}}\sqrt{\frac{8\pi}{k}}\begin{Bmatrix} Y_o \\ Z_o\end{Bmatrix}\hat{e} \\ &\qquad\cdot\begin{Bmatrix} \overline{E}^i(Q_E)D_{es}^U(\phi,\phi';\sqrt{\sin\beta_o\sin\beta_o'}) \\ \overline{H}^i(Q_E)D_{eh}^U(\phi,\phi';\sqrt{\sin\beta_o\sin\beta_o'})\end{Bmatrix}.\end{aligned} \tag{58}$$

Recently, a new formulation of the PTD was presented in [20] for directly calculating the fringe wave contribution pertaining to the scalar (acoustic) case. Those ideas in [20] can be directly extended here to find $\overline{E}_U^S$ for the vector EM case; thus

$$\overline{E}_U^S(P) \sim \frac{-jkZ_o}{4\pi}\int\int_S \overline{J}^d(\bar{r}')\cdot\left[\overline{\overline{I}} + \frac{\nabla\nabla}{k^2}\right]\frac{e^{-jk|\bar{r}-\bar{r}'|}}{|\bar{r}-\bar{r}'|}ds' \tag{59}$$

since $\overline{E}_U^S$ is radiated by the component of the current which is produced by the edge diffracted field on the surface of the scatterer; this diffracted component of the current is denoted by $\overline{J}^d(\bar{r}')$ in (59). According to [20],

$$ds' = \frac{d\tau d\sigma}{|J_{\tau\sigma}|} \tag{60}$$

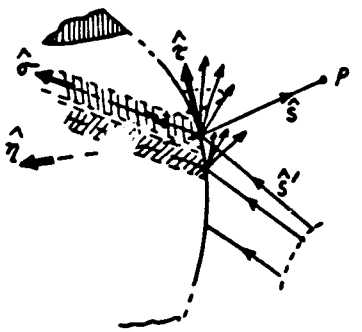
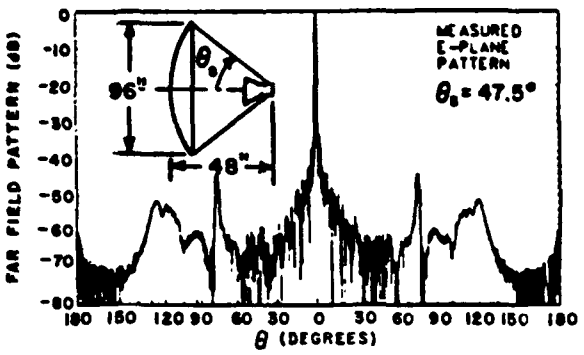


Fig. 22. Integration coordinates on the wedge.

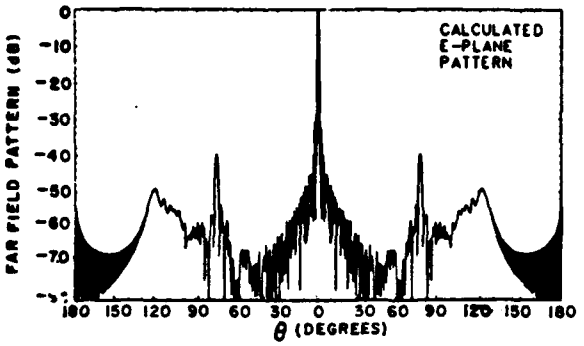
where the coordinate σ is along the Keller cone of diffracted rays on the surface. The coordinates τ, σ and η are shown in Fig. 22; it is noted that $\hat{\tau} = \hat{e}$ = unit edge tangent vector, and $\hat{\eta} \perp \hat{e}$. The integral in (59) together with (60) may be evaluated asymptotically in closed form along the σ coordinate, thereby leaving a line integral along the τ variable (i.e., along the edge contour as in (57)) that yields a PTD based ECM interpretation (analogous to (57)) from which a more refined set of equivalent currents I^U and M^U than those in (58) can be identified. It is noted that only the dominant range dependent terms may be retained in (52) and (59) which result from the $\nabla\nabla$ operation therein; the remaining higher order range terms may be neglected as usual for $k|\bar{r}-\bar{r}'| \gg 1$.

The PTD can also be employed to deal with apertures. In this case, the PO concepts may be extended so that the PO type contribution can be found from the GO fields in the aperture; such a PO integral over the aperture is commonly referred to as the aperture integral (AI) when it is applied to horn and reflector antennas. An appropriate Ufimtsev correction $\overline{E}_U^S$ may then be added to $\overline{E}_{AI}^S$ (corresponding to the AI contribution which acts like the PO contribution). It is noted that $\overline{E}_{PO}^s$, or $\overline{E}_{AI}^S$ for aperture problems, is generally far more significant than the Ufimtsev correction $\overline{E}_U^S$ in the region corresponding to the main beam, as for example in the case of horn and reflector antennas.

Figure 23 illustrates the far zone E plane radiation pattern of a symmetric parabolic reflector fed at the focus, with four symmetric struts holding the feed; this pattern has been calculated in [58] using AI up to 6° away from the main beam axis and switching to UTD beyond 6° (except for the use of GTD-based ECM to patch up GTD in the rear axial caustic direction at $\theta = 180°$). The Ufimtsev correction to AI is negligible in Fig. 23 which shows that the AI alone compares very well with measurements. The diameter of the conducting struts is 0.84 in and the scattering from these struts is found by using a wire diffraction coefficient in conjunction with ECM as described in [58]. Figure 24 shows the far zone E plane radiation pattern of a pyramidal horn calculated in [59] using the AI technique to 30° away from the main beam axis, and the GTD-based ECM beyond 30°. Measured results are also shown for comparison in Fig. 24 from which it can be again seen that the Ufimtsev correction to AI is negligible in this case. The Ufimtsev correction to



(a)



(b)

Fig. 23 Measured and calculated E plane patterns of a symmetric parabolic reflector antenna at 11 GHz. (a) Measured. (b) Calculated (see [58]).

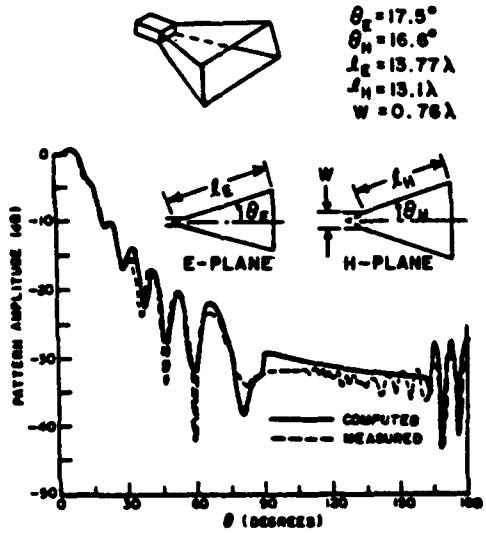


Fig. 24 Measured and calculated E plane patterns of a pyramidal horn antenna (see [59]).

AI becomes important if one employs AI not just to calculate the antenna main beam and the first few sidelobes, but to also calculate the wide angle sidelobes and back lobes.

III. Conclusions

It is seen that high-frequency techniques are conceptually simple as well as versatile in being able to predict the radiation patterns, mutual coupling and other effects associated with a large variety of practical antenna configurations. However, as also seen from Section II, the use of GTD/UTD technique requires a knowledge of the relevant diffraction coefficients; therefore, while several diffraction mechanisms can presently be characterized by UTD coefficients, more UTD coefficients need to be developed to solve a greater variety of antenna problems which are relevant to present and future EM technology. Some UTD coefficients which are known only approximately at present need to be refined in some cases; others need to be found for additional perfectly conducting as well as nonconducting (and even penetrable) canonical structures. Some work in the latter case which is available in [60], [61] needs to be developed further; such work would be useful, for example, to predict the reduction in coupling between antennas on a metal surface by introducing a lossy (absorbing) material patch placed between the pair of antennas, or to predict the effects of the canopy of private aircraft, or helicopters, on the antennas located on such structures, etc. The PTD likewise needs to be formally extended to deal with nonconducting surfaces and to smooth surfaces without edges, etc. Not discussed in this paper are spectral techniques which can be used in conjunction with high-frequency approximations to deal with complex (nonray optical) illumination of the host structure by the primary source (antenna) [62], [63]. Finally, hybrid procedures which combine high and low frequency techniques [66], and the Gaussian beam techniques have also not been discussed here due to space limitations. Different aspects of ray and Gaussian beam methods have appeared previously as a collection of papers in [65]; the latter also contains a paper by Borovikov and Kinber, which in turn provides a large bibliography of Soviet papers on high-frequency techniques. The hybrid procedures as well as the Gaussian beam techniques appear to hold promise to solve some high-frequency EM antenna and scattering problems which may otherwise become intractable.

References

[1] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, pp. 116–130, 1962.

[2] R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, vol. 62, pp. 1448–1461, Nov. 1974.

[3] ——, "A uniform GTD approach to EM scattering and radiation," in *Acoustic, Electromagnetic and Elastic Wave Scattering —High and Low Frequency Asymptotics*, vol. II, Varadan and Varadan, Eds. Amsterdam, The Netherlands: North Holland, 1986.

[4] P. H. Pathak, "Techniques for high frequency problems," in *Antenna Handbook, Theory Application and Design*, Y. T. Lo and S. W. Lee, Eds. New York: Van Nostrand Reinhold, 1988.

[5] S. W. Lee and G. A. Deschamps, "A Uniform Asymptotic theory of EM diffraction by a curved wedge,"*IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 25–34, Jan. 1976.

[6] R.C. Hansen, Ed., *Geometrical Theory of Diffraction.* New York: IEEE Press, 1981.

[7] G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves.* IEE EM Wave Series 1, (Peter Peregrinus), 1976.

[8] D. A. McNamara, C. W. I. Pistorius, and J. A. G. Malherbe, *Introduction to the Uniform Geometrical Theory of Diffraction.* New York: Artech House, 1990.

[9] C. A. Balanis, *Advanced Engineering Electromagnetics*, IEE EM Wave Series 10, (Peter Peregrinus), 19.
[10] E. V. Jull, *Aperture Antennas and Diffraction Theory*, IEE EM Wave Series 10, (Peter Peregrinus), 1981.
[11] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation and Scattering*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
[12] V. A. Fock, *Electromagnetic Diffraction and Propagation Problems*. New York: Pergamon, 1965.
[13] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
[14] P. H. Pathak and M. C. Liang, "On a uniform asymptotic solution valid across smooth caustics of rays reflected by smoothly indented boundaries," *IEEE Trans. Antennas Propagat.*, vol. AP-38, pp. 1192–1203, Aug. 1990.
[15] J. J. Stamnes, *Waves in Focal Regions*. Bristol and Boston, MA: Adam Higler 1986.
[16] C. E. Ryan Jr. and L. Peters Jr., "Evaluation of edge diffracted fields including equivalent currents for caustic regions," *IEEE Trans. Antennas Propagat.*, vol. AP-7, pp. 292–299, May 1969; also, see correction, *IEEE Trans. Antennas Propagat.*, vol. AP-8, p. 275, Mar. 1970.
[17] E. F. Knott and T. B. A. Senior, "Comparison of three high-frequency diffraction techniques," *Proc. IEEE*, vol. 62, pp 1468–1478, Nov. 1974.
[18] R. F. Harrington, *Time Harmonic Electromagnetic Fields*. New York: McGraw-Hill, 1961.
[19] P. Ya Ufimtsev, "Method of edge waves in the physical theory of diffraction," (from the Russian "Method Krayevykh Voin V Fizicheskoy Theorii Difraktsii," *Izd-Vo Sov. Radio*, pp. 1–243, 1962), Translation prepared by the U. S. Air Force Foreign Technology Division, Wright-Patterson AFB, OH; released for public distribution Sept. 7, 1971.
[20] D. I. Butorin and P. Ya Ufimtsev, *Sov. Phys. Acoust.*, vol. 32, no. 4, July–Aug. 1986.
[21] M. S. Narasimhan and K. M. Prasad, "GTD analysis of near field patterns of a prime focus symmetric parabolic reflector antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 959–961, Nov. 1981.
[22] A. Michaeli, "Transition functions for high frequency diffraction by a curved perfectly conducting wedge—parts I, II and III," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 1073–1092, Sept. 1989.
[23] M. C. Liang, C. W. Chuang, and P. H. Pathak, Paper on edge diffracted surface rays, in preparation.
[24] M. C. Liang, P. H. Pathak, and C. W. Chuang, "A generalized UTD analysis for the diffraction by a wedge with convex faces to include surface ray effects and grazing angles of incidence/diffraction," presented at the 1988 Int. IEEE APS/Radio Science Meeting, Syracuse Univ., Syracuse, NY, June 6–10, 1988.
[25] M. C. Liang, "A generalized uniform GTD ray solution for the diffraction by a perfectly-conducting wedge with convex faces," Ph.D. dissertation, The Ohio State Univ., 1988.
[26] M. Idemen and L. B. Felsen, "Diffraction of a whispering gallery mode by the edge of a thin concave cylindrically curved surface," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 571–579, July 1981.
[27] K. C. Hill and P. H. Pathak, "On the computation of the transition function occurring in a new approximate UTD corner diffraction coefficient," in preparation.
[28] K. C. Hill, "A UTD solution to the EM scattering by the vertex of a perfectly-conducting plane angular sector," Ph.D. dissertation, The Ohio State Univ., 1990.
[29] F. A. Sitka, W. D. Burnside, T. T. Chu, and L. Peters Jr., "First order equivalent current and corner diffraction scattering from flat plate structures," *IEEE Trans. Antennas Propagat.*, vol. AP-31, pp 584–589, July 1983.
[30] K. E. Golden, G. E. Stewart, and D. C. Pridmore-Brown, "Approximation techniques for the mutual admittance of slot antennas on metallic cones," *IEEE Trans. Antennas Propagat.*, vol. AP-22, pp. 43–48, 1974.
[31] K. D. Trott, P. H. Pathak, and F. A. Molinet, " A UTD type analysis of the plane wave scattering by a fully Illuminated perfectly conducting cone," *IEEE Trans. Antennas Propagat.*, vol. AP-38, pp. 1150–1160, Aug. 1990.
[32] P. H. Pathak, "An asymptotic analysis of the scattering of plane waves by a smooth convex cylinder," *Radio Sci.*, Vol. 14, pp. 419–435, May–June 1979.
[33] P. H. Pathak, W. D. Burnside, and R. J. Marhefka, "A uniform UTD analysis of the diffraction of electromagnetic waves by a smooth convex surface," *IEEE Trans. Antennas Propagat.*, vol. AP-28, pp. 609–622, Sept. 1980.
[34] N. A. Logan and K. S. Yee, "A mathematical model for diffraction by convex surfaces," in *Electromagnetic Waves*, R. Langer, Ed., Madison, WI: Univ. Wisconsin Press, 1962.
[35] B. R. Levy and J. B. Keller, "Diffraction by a smooth object," *Comm. Pure Appl. Math.*, vol. 12, pp. 159–209, Feb. 1959.
[36] P. H. Pathak, N. Wang, W. D. Burnside and R. G. Kouyoumjian, "A uniform GTD solution for the radiation from sources on a convex surface," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 609–621, July 1981.
[37] P. H. Pathak and N. Wang, "Ray analysis of mutual coupling between antennas on a convex surface," *IEEE Trans. Antennas Propagat.*, vol. AP-29, Nov. 1981.
[38] S. W. Lee, "Mutual admittance of slots on a cone: Solution by ray technique," *IEEE Trans. Antennas Propagat.*, vol. AP-26, pp. 768–773, Nov. 1978.
[39] J. J. Kim and W. D. Burnside, "Simulation and analysis of antennas radiating in a complex environment," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 554–562, Apr. 1986.
[40] P. H. Pathak, A. Ghantous, C. W. Chuang, and O. M. Buyukdura, "A preliminary investigation of conformal antenna phased arrays on perfectly-conducting convex surfaces," The Ohio State Univ. ElectroScience Lab., Tech. Rep. 717060-2, (prepared under Contract L5XM-379206-947, for Rockwell International, Los Angeles, CA,) Nov. 1985.
[41] H. Ikuno and L. B. Felsen, "Complex ray interpretation of reflection from concave–convex surfaces," *IEEE Trans. Antennas Propagat.*, vol. 36, pp. 1206–1271, Sept. 1988.
[42] R. F. Millar, "An approximate theory of the diffraction of an electromagnetic wave by an aperture in a plane screen," *Proc. Inst. Elect. Eng.*, vol. 103C, pp. 177–185, 1956.
[43] C. R. Cockrell and P. H. Pathak, "Diffraction theory techniques applied to aperture antennas on finite circular and square ground planes," *IEEE Trans. Antennas Propagat.*, vol. AP-22, pp. 443–448, May 1974.
[44] C. A. Balanis, "Radiation from conical surfaces used for high speed aircraft," *Radio Sci.*, vol. 7, pp. 339–343, Feb. 1972.
[45] H. Y. Yee and L. B. Felsen, "Ray optical techniques for waveguide discontinuities," Rep. PIBEP-68-005, Polytechnic Inst. of Brooklyn, NY, 1968.
[46] H. Y. Yee, L. B. Felsen, and J. B. Keller, "Ray theory of reflection from the open end of a waveguide," *SIAM J. Appl. Math.*, vol. 16, pp. 268–300, 1968.
[47] S. W. Lee, "Ray theory of diffraction by open-ended waveguides: Applications," *J. Math. Phys.*, vol. 13, pp. 656–664, 1972.
[48] J. Boersma, "Ray-optical Analysis of Reflections in an open ended parallel plane Waveguide: TM case," *SIAM J. Appl. Math.*, vol. 29, pp. 164–195, 1975.
[49] R. C. Rudduck and L. L. Tsai, "Aperture reflection coefficients for TEM and TE_{10} mode parallel-plate waveguides," *IEEE Trans. Antennas Propagat.*, vol. AP-16, pp. 83–89, 1968.
[50] P. H. Pathak and A. Altintas, "An efficient high-frequency analysis of modal reflection and transmission coefficients for a class of waveguide discontinuities," *Radio Sci.*, vol. 23, no. 6, pp. 1107–1119, Nov.–Dec. 1988.
[51] C. W. Chuang and P. H. Pathak, "Ray analysis of modal reflection for three-dimensional open-ended waveguides," *IEEE Trans. Antennas and Propagat.*, vol. 37, pp. 339–346, Mar. 1989.
[52] A. Michaeli, "Elimination of infinities in equivalent edge currents, parts I and II," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 912–918, July 1986 and pp. 1034–1037, Aug. 1986.
[53] K.M. Mitzner, "Incremental length diffraction coefficients," Aircraft Div., Northrop Corp., Tech. Rep. AFAL-TR-73-296, Apr. 1974.
[54] E.F. Knott, "The relationship between Mitzner's ILDC and Michaeli's equivalent currents," *IEEE Trans. Antennas Propagat.*, vol. AP-33, pp. 112–114, Jan. 1985.
[55] R. A. Shore and A. D. Yaghjian, "Incremental diffraction coefficients for planar surfaces," *IEEE Trans. Antennas Propagat.*, vol. 36, pp. 55–70, Jan. 1988.
[56] I. J. Gupta, C. W. I. Pistorius, and W. D. Burnside, "An efficient method to compute spurious endpoint contributions in PO solutions," *IEEE Trans. Antennas Propagat.*, vol. AP-35, Dec. 1987.

[57] S. W. Lee, "Comparison of uniform asymptotic theory and Ufimtsev's theory of electromagnetic edge diffraction," *IEEE Trans. Antennas Propagat.*, vol. AP-25, pp. 162–170, Mar. 1977.
[58] T. H. Lee, R. C. Rudduck, and K. M. Lambert, "Pattern measurements of Reflector Antennas in the compact rand and validation with computer code simulation," *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 889–895, June 1990.
[59] C. A. Mentzer, L. Peters Jr., and R. C. Rudduck, "Slope diffraction and its application to horns," *IEEE Trans. Antennas Propagat.*, vol. AP-23, pp. 153–159, Mar. 1975.
[60] R. G. Rojas and P. H. Pathak, "Diffraction of EM waves by a dielectric/ferrite half-plane and related configurations," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 751–763, June 1989.
[61] J. L. Volakis and T. B. A. Senior, "Diffraction by a thin dielectric half plane," *IEEE Trans. Antennas Propagat.*, vol. AP-35, pp. 1483–1487, Dec. 1987.
[62] Y. Rahmat-Samii and R. Mittra, "A spectral domain interpretation of high frequency phenomena," *IEEE Trans. Antennas Propagat.*, vol. AP-25, pp. 676–687, Sept. 1977.
[63] R. Tiberio, G. Manara, G. Pelosi, and R.G. Kouyoumjian, "High-frequency electromagnetic scattering of plane waves from double wedges," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 1172–1180, Sept. 1989.
[64] E. K. Miller and G. J. Burke, "Low frequency computational electromagnetics for antenna analysis," *Proc. IEEE*, vol. 80, this issue.
[65] G. A. Thiele, "Overview of selected hybrid methods in radiating system analysis," *Proc. IEEE*, vol. 80, this issue.
[66] Special Issue on Rays and Beams, *Proc. IEEE*, , vol. 62, Nov. 1974.

Prabhakar H. Pathak (Fellow, IEEE) received the B.Sc. degree in physics form the University of Bombay, India, in 1952, the B.S. degree in electrical engineering from the Louisiana State University, Baton Rouge, in 1965, and the M.S. and Ph.D. degrees from The Ohio State University (OSU), Columbus, in 1970 and 1973, respectively.

From 1965 to 1966 he was an Instructor in the Department of Electrical Engineering at the University of Mississippi, Oxford. During the summer of 1966 he worked as an electronics engineer with Boeing Company, Renton, WA. Since 1963 he has been with the OSU Electro-Science Laboratory. His research intersets are in mathematical methods, electromagnetics, and uniform ray techniques. In 1983, he joined the faculty of the Department of Electrical Engineering, The Ohio State University, where he is currently a Professor. He has contributed to the development of the uniform geometrical theory of diffraction (UTD), which can be applied to analyze a variety of practical electromagnetics antenna and scattering problems. His work continues to be in the UTD and other high-frequency techniques, as well in the analysis of guided waves, and microstrip and reflector antennas using high-frequency asymptotic procedures. He has presented invited lectures, and several short courses on the UTD, both in the U.S. and abroad. He has also authored and coauthored chapters on ray methods for five books. He has served as Associate Editor for the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION for about five years.

Dr. Pathak is a member of Commission B of the International Scientific Radio Union (URSI) and Sigma Xi.

On The Dyadic Green's Function For a Planar Multilayered Dielectric/Magnetic Media

Sina Barkeshli, *Member, IEEE*, and P. H. Pathak, *Fellow, IEEE*

***Abstract*—A complete plane wave spectral eigenfunction expansion of the electric dyadic Green's function for a planar multilayered dielectric/magnetic media is given in terms of a pair of the ($\hat{z}$)-propagating solenoidal eigenfunctions, where ($\hat{z}$) is normal to the interface, and it is developed via a utilization of the Lorentz reciprocity theorem. This expansion also contains an explicit dyadic delta function term which is required for completeness at the source point. Some useful concepts such as the effective plane wave reflection and transmission coefficients are employed in the present spectral domain eigenfunction expansion. The salient features of this Green's function are also described along with a physical interpretation.**

I. Introduction

A COMPLETE plane wave spectral (PWS) type eigenfunction expansion of the electric dyadic Green's function for the planar multilayered dielectric/magnetic media is given in this paper in terms of a pair of the ($\hat{z}$)-directed solenoidal eigenfunctions, where ($\hat{z}$) is normal to the interface, and it is developed via a utilization of the Lorentz reciprocity theorem. This expansion also contains an explicit dyadic delta function term which is required for making the representation complete at the source point. The geometry of this problem is shown in Fig. 1. The electrical parameters in each of the layers are assumed to be homogeneous and isotropic. It is shown that the field at a given point consists of four distinct wave types (two for each TE and TM type) caused by the presence of the multilayered media. This dyadic Green's function is useful in many problems dealing with the stratified media, i.e., scattering from buried objects in the layered earth, or in the design of high performance finite phased arrays in multilayered dielectric/magnetic environment. Since the dyadic Green's function derived here is for an arbitrarily oriented current point source, it can also be utilized for the applications where the current elements are obliquely rather than horizontally or vertically oriented with respect to the planar interfaces.

The plane wave spectrum (PWS) integral representation of the dyadic Green's function for this canonical problem may be constructed in several ways. One of the most common approaches is to express the Green's function in terms of a magnetic vector potential [1]–[5], whereas another approach is to construct the Green's function from a set of appropriate electric and magnetic vector potentials [6]–[10], [21]. In the former case, the magnetic vector potential in general has components which are parallel and normal to the interface even if the electric point current source does not possess a component which is normal to the interface. In the other approach, the magnetic and electric vector potentials are generally chosen so that they are both normal to the interface. If the electric point current source is chosen normal to the interface, then the two approaches become identical since only a single normally directed magnetic vector potential suffices in this case. This is related to the fact that the normally oriented current moment only excites the TM waves (with respect to the ($\hat{z}$)-coordinate direction), whereas the electric current moment parallel to

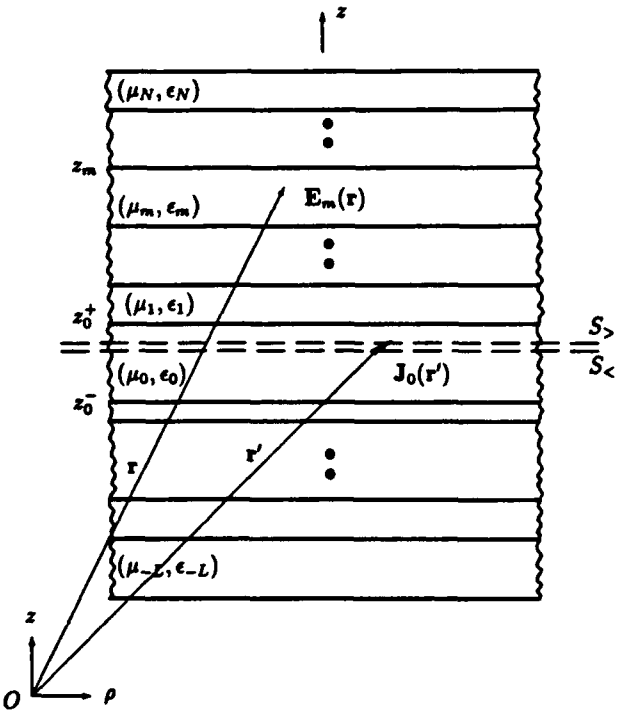


Fig. 1. Electric point current dipole source in a multi-layered dielectric/magnetic media. Also the planar surfaces $S_>$ and $S_<$ slightly above and below the source are shown.

Manuscript received March 7, 1990; revised May 29, 1991. This work was supported in part by the Joint Services Electronics Program, Contract N00014-78-C-0049 and the Ohio State University Research Foundation.

S. Barkeshli is with Sabbagh Associates, Inc., 4639 Morningside Drive, Bloomington, IN 47401.

P. H. Pathak is with the ElectroScience Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, OH 43212.

IEEE Log Number 9102812.

the interface excites both TM and TE waves. Therefore the total electromagnetic waves must be constructed either with the magnetic vector potential which can produce both TM and TE waves (in this case magnetic vector potential must have components normal and parallel to the interface) in order to satisfy the appropriate boundary conditions, or with the magnetic and electric vector potentials which are both normal to the interface (since a normally directed magnetic vector potential produces TM waves and a normally directed electric vector potential produces TE waves). One of the main advantages of the latter formulation is that the boundary conditions associated with the differential operators for the two different types of vector potentials can be decoupled. In the case of a choice of a single type of magnetic vector potential containing both a vertical ($\hat{z}$) and a horizontal (transverse) to ($\hat{z}$) component, the transverse component (parallel to interface) of that magnetic vector potential will contribute to both TE and TM waves; therefore, the boundary conditions for normal and transverse potential components will be coupled. This disadvantage will be more pronounced if one deals with the stratified or multilayer dielectric/magnetic media, for which the number of coupled boundary conditions increase, thereby complicating the analysis. Recently Bagby and Nyquist [11], derived a formal representation of the dyadic Green's function for the multilayered media in terms of the magnetic vector potential [1], [4], which they specialized for the cases of microstrip and optical circuit structures. Since only the magnetic vector potential is used, the boundary conditions for the TM and TE waves are coupled in [11], hence, the natural distinction between the two is lost. Also the dyadic delta function term, which makes the representation complete at the source point, was not explicitly extracted in [11]; Viola and Nyquist [12], slightly modified that analysis later to properly extract the dyadic delta function term. In the present work, we have derived a complete eigenfunction expansion of the dyadic Green's function for the planar multilayered dielectric/magnetic media using the ($\hat{z}$)-directed solenoidal electric and magnetic (TM and TE) eigenfunctions. We have used continuous eigenmodes propagating along a "preferred" ($\hat{z}$)-direction. We have also employed the orthogonality properties of the eigenmodes over an open planar surface [6] transverse to the direction of the propagation, ($\hat{z}$) to construct our Green's dyadic. This is a generalization of the discrete eigenvalues and eigenmodes, that is usually used in the guided wave theory [13]. Hence, unlike the work reported previously, this analysis retains the connection between the closed (waveguides) and open (planar multilayer) type structure, which is usually lost in the formal Fourier transform method. In addition, because those eigenvalues and eigenmodes are only a function of the geometry of structure, and not the excitation [6], [13], the natural (TM and TE) eigenmodes reveal the physical behavior of the fields in the multi-layered dielectric/magnetic media. Finally, we have employed a method that utilizes only the solenoidal eigenfunctions [14], and hence, the dyadic delta function term at the source point is included explicitly as a correction to the general solenoidal eigenfunction expansion which is valid outside the source point. The electric dyadic Green's components given in this work appear to be closely related to those electric field components which have been derived by Kong [7], [8], and Chew [21] utilizing the usual boundary conditions at each of the interfaces and the proper condition at the source point. As indicated above, the procedure used here is somewhat different, in that we have utilized the orthogonality of continuous eigenmodes at the planar interfaces along with the Lorentz reciprocity theorem to drive the complete eigenfunction expansion of the electric dyadic Green's function which contains a physical interpretation.

The format of the paper is as follows. In Section II, we outline the procedure required to derive the complete eigenfunction expansion of the dyadic Green's function for the multilayered media, $\mathcal{G}^{m,0}$, in terms of only the solenoidal eigenfunctions. In Section III, we start with the unbounded case, in which the point source radiates with no interface present, and construct the corresponding dyadic Green's function, $\mathcal{G}^0$, in terms of an integral over the spectra of plane waves that constitute the continuous eigenfunction expansion in which the eigenfunctions are guided in the preferred $\hat{z}$-coordinate direction, using the procedure described in Section II. This is essentially the z-propagation (plane wave spectrum) representation of the free space dyadic Green's function which is usually represented by the discrete spherical vector wave type radially propagating eigenfunction expansion. In Section IV, the dyadic Green's function for the multilayered media, $\mathcal{G}^{m,0}$, is then constructed from the principle of the superposition, which involves the sum of the fields of firstly the source in free space (or the free space Green's function $\mathcal{G}^0$) and secondly the fields scattered by the layered media. Section V deals with the physical interpretation of the dyadic Green's function and numerical results. Conclusions and discussions are presented in Section VI.

II. Formulation of $\mathcal{G}^{m,0}$ in Terms of the Solenoidal Eigenfunctions

In this section we outline a general procedure described by Pathak, [14], which can also be employed to find a complete eigenfunction expansion of the electric field in the multilayered media, $\boldsymbol{E}_m$, and its corresponding dyadic analog $\mathcal{G}^{m,0}$ in terms of only the solenoidal eigenfunctions.

The usual Maxwell curl equations for the electric and magnetic fields $\boldsymbol{E}_m$ and $\boldsymbol{H}_m$ within any mth layer (see Fig. 1), respectively, are given by

$$\nabla \times \boldsymbol{E}_m = -j\omega\mu_m \boldsymbol{H}_m; \quad \nabla \times \boldsymbol{H}_m = j\omega\epsilon_m \boldsymbol{E}_m + \boldsymbol{J}_0. \quad (1)$$

An $e^{j\omega t}$ time dependence is assumed and suppressed in (1), and as usual, μ_m and ϵ_m are the permeability and permittivity of the medium (m), and $\boldsymbol{J}_0$ is the impressed electric current source. If the electric current density $\boldsymbol{J}_0$ is

taken to be a point source of strength p_e at $r = r'$ in the region (0); then,

$$J_0(r) = p_e\,\delta(r - r'). \tag{2}$$

Before proceeding further, it is important to relate the dyadic Green's function to the electric field due to J_0 as [15]

$$E_m(r) = -j\omega\mu_0 \iiint_v \mathcal{G}^{m,0}(r, r'') \cdot J_0(r'')\,dv'', \tag{3}$$

where $\mathcal{G}^{m,0}$ is the multilayered electric dyadic Green's function, and v contains the source region. If $J_0(r'')$ is an arbitrarily oriented point current source of the strength p_e given in (2), then electric field may be viewed as a distribution; namely,

$$E_m(r) = -j\omega\mu_0 \mathcal{G}^{m,0}(r, r') \cdot p_e. \tag{4}$$

Let the solenoidal part of the eigenfunction expansion of the electric field E_m, which is valid for $z \neq z'$ (and hence for $r \neq r'$), be denoted by E'_m. The field E'_m is obtained in terms of only the solenoidal eigenfunctions because the electric field has zero divergence for $z \neq z'$. The z-propagating solenoidal eigenfunction expansion of E'_m can be expressed as

$$E'_m = \begin{cases} E_m^{>}, & z > z' \\ E_m^{<}, & z < z'. \end{cases} \tag{5}$$

Alternatively, E'_m in (5) can be written as

$$E'_m = \mathcal{U}(z - z')E_m^{>} + \mathcal{U}(z' - z)E_m^{<}, \tag{6}$$

where the Heaviside unit step function $\mathcal{U}(\xi)$ is defined by,

$$\mathcal{U}(\xi) = \begin{cases} 1, & \xi > 0 \\ 0, & \xi < 0 \end{cases},$$

and $\gtrless$ means the fields for $z \gtrless z'$. The entire space consists of two regions $z > z'$ and $z < z'$; $z = z'$ is the plane S (normal to $\hat{z}$-axis) containing the source, $J_0 = p_e\,\delta(\rho - \rho')\,\delta(z - z')$, in region (0) of Fig. 1. It is noted that (μ_0, ϵ_0) correspond to the constitutive parameters of the medium in region (0); in general, (μ_0, ϵ_0) are different from those for free space. Consider next the magnetic field H_m due to J_0; in particular making use of (5), yields

$$\nabla \times E_m^{\gtrless} = -j\omega\mu_m H_m^{\gtrless}; \quad \nabla \times H_m^{\gtrless} = j\omega\epsilon_m E_m^{\gtrless}, \tag{7}$$

where $H_m^{\gtrless}$ is the value of the magnetic field H_m in the region (m), for $z \gtrless z'$. It is clear from (7) that the magnetic field $H_m^{\gtrless}$ is known once $E_m^{\gtrless}$ is known. The fields $H_m^{>}$ and $H_m^{<}$ must satisfy the proper source condition at $r = r'$. In order to impose the boundary condition at the source point, $r = r'$, the volume current density J_0 must be expressed in terms of a distribution p_{es} corresponding to a "surface" current density at $z = z'$ (i.e., on the surface S); thus

$$J_0 = p_e\,\delta(\rho - \rho')\,\delta(z - z') = P_e\,\delta(r - r')$$
$$= p_{es}\,\delta(z - z'). \tag{8}$$

Now the discontinuity of the tangential magnetic field in the region (0), across S (at $z = z'$) must be equal to the surface current density at S; namely,

$$\hat{z} \times (H_0^{>} - H_0^{<}) = \hat{I}_t \cdot p_{es}, \tag{9}$$

where $\hat{I}_t$ denotes the transverse part of the unit dyad with respect to $\hat{z}$,

$$\hat{I} = \hat{I}_t + \hat{z}\hat{z}; \quad \hat{I}_t = \hat{x}\hat{x} + \hat{y}\hat{y}. \tag{10}$$

It is clear that (9) is valid only at $z = z'$, so it can be expressed as

$$\hat{z} \times (H_0^{>} - H_0^{<})\,\delta(z - z') = \hat{I}_t \cdot p_{es}\,\delta(z - z'), \tag{11}$$

it follows directly from (8) that the above equation becomes

$$\hat{z} \times (H_0^{>} - H_0^{<})\,\delta(z - z') = \hat{I}_t \cdot p_e\,\delta(r - r'),$$

or more generally,

$$\hat{z} \times (H_m^{>} - H_m^{<})\,\delta(z - z') = \hat{I}_t \cdot p_e\,\delta(r - r'). \tag{12}$$

This is the expression for the condition on H_m at the source point, and it directly indicates the appropriate addition to E'_m at the source point which is required to yield the complete expansion of E_m. It is important to note that since the discontinuity condition in (12) relates $H_m^{>}$ to $H_m^{<}$ across the source point, one only needs to know $H_m^{>}$ and $H_m^{<}$ to completely specify H_m due to the source $J_0 = p_e\,\delta(r - r')$; thus

$$H_m = \mathcal{U}(z - z')H_m^{>} + \mathcal{U}(z' - z)H_m^{<}. \tag{13}$$

The E'_m of (6) can now be readily found by employing (7), and using the relation based on distribution theory, [16],

$$\nabla \times [H_m^{\gtrless}\,\mathcal{U}(\pm z \mp z')] = \mathcal{U}(\pm z \mp z')\nabla \times H_m^{\gtrless}$$
$$\pm \hat{z} \times H_m^{\gtrless}\,\delta(z - z'). \tag{14}$$

From (7), (12) and (14), if follows that

$$E'_m = \frac{1}{j\omega\epsilon_m}\nabla \times H_m - \frac{1}{j\omega\epsilon_0}\hat{I}_t \cdot p_e\,\delta(r - r'). \tag{15}$$

The precise relationship between the complete field E_m of (4) and the incomplete field E'_m of (15) can now be written by using (1) and (10), [14],

$$E_m(r, r') = E'_m(r, r') - j\omega\mu_0\left[-\frac{\hat{z}\hat{z}}{k_0^2}\,\delta(r - r')\right] \cdot p_e;$$
$$E_{m|z=z'} = E'_{m|z=z'}. \tag{16}$$

The Green's dyadic $\mathcal{G}^{m,0}$ can be inferred from (16) by comparison with (4). Thus,

$$\mathcal{G}^{m,0}(r, r') = \bar{\bar{g}}^{m,0}(r, r') - \frac{\hat{z}\hat{z}}{k_0^2}\,\delta(r - r'), \tag{17}$$

with

$$E'_m(r, r') \equiv -j\omega\mu_0 \bar{\bar{g}}^{m,0}(r, r') \cdot p_e. \quad (18)$$

From the above discussion, it is clear that one can also construct the complete free space dyadic Green's function, $\mathcal{G}^0$, in terms of the "z-propagating" solenoidal eigenfunctions which will be obtained in the following section.

III. Construction of the Free Space Dyadic Green's Function, $\mathcal{G}^0$

In this section, the procedure outlined in the previous section is applied to obtain an explicit expansion for $\mathcal{G}^0$ which is associated with an electric point current source, $J_0 = p_e\,\delta(r - r')$, which radiates in an unbounded medium with parameter (μ_0, ϵ_0) which are the same as in region (0) with no interface present. In the following section, the procedure developed here will be extended to explicitly obtain the dyadic Green's function $\mathcal{G}^{m,0}$ of (17) for the multilayered media. The first step in the procedure for obtaining the free space electric field E_0 and its corresponding $\mathcal{G}^0$ involves the construction of a z-propagating PWS solenoidal eigenfunction expansion of E'_0 which is complete if $z \neq z'$.

The geometry of the problem dealing with a homogeneous (free) space with constitutive parameters (μ_0, ϵ_0) excited by $J_0 = p_e\,\delta(r - r')$ is illustrated in Fig. 2. The solenoidal eigenfunctions for this problem are chosen to propagate in the preferred $\pm\hat{z}$-coordinate direction. The source point at $r = r'$ lies in the plane S at $z = z'$ as in Fig. 2. Let $E^{\gtrless}$ and $H^{\gtrless}$ denote the continuous PWS solenoidal vector wave function expansions for the electric and magnetic fields, due to J_0 in the absence of the interface; thus,

$$E^{\gtrless} = E'^{\gtrless} + E''^{\gtrless}; \quad H^{\gtrless} = H'^{\gtrless} + H''^{\gtrless}, \quad (19)$$

and [6],

$$E^{\gtrless} = \int dk_t\,(a'^{\gtrless} e'^{\gtrless} + a''^{\gtrless} e''^{\gtrless});$$

$$H^{\gtrless} = \int dk_t\,(a'^{\gtrless} h'^{\gtrless} + a''^{\gtrless} h''^{\gtrless}), \quad (20)$$

where prime (') and double prime (") refer to TM and TE wave components with respect to the preferred $\hat{z}$-coordinate direction, respectively, and h' and e'' can be derived from the solenoidal magnetic and electric $\hat{z}$-directed vector potentials, (Π', Π''), respectively [6], [13],

$$h'^{\gtrless} = \mp\hat{z} \times \nabla_t \Pi'^{\gtrless};$$

$$e'^{\gtrless} = \pm\frac{1}{j\omega\epsilon_0}\nabla_t \frac{\partial}{\partial z}\Pi'^{\gtrless} \mp \frac{1}{j\omega\epsilon_0}\hat{z}\nabla_t^2\Pi'^{\gtrless}, \quad (21)$$

and

$$e''^{\gtrless} = \hat{z} \times \nabla_t \Pi''^{\gtrless};$$

$$h''^{\gtrless} = \frac{1}{j\omega\mu_0}\nabla_t \frac{\partial}{\partial z}\Pi''^{\gtrless} - \frac{1}{j\omega\mu_0}\hat{z}\nabla_t^2\Pi''^{\gtrless}, \quad (22)$$

where, ∇_t is the transverse (to $\hat{z}$) part of the ∇ operator. Electric and magnetic potentials, Π' and Π'', which can also be viewed as a pair of Debye potentials [5], [17] satisfy the well known Helmholtz equation; their associated $\hat{z}$-propagating eigenfunctions can be expressed as

$$\Pi^{\gtrless}(k_t) = \frac{1}{2\pi}\exp[-j(k_t \cdot \bar{\rho} \pm \kappa_0 z)];$$

$$\Pi^{\gtrless}(-k_t) = \frac{1}{2\pi}\exp[-j(-k_t \cdot \bar{\rho} \pm \kappa_0 z)]; \quad (23)$$

where prime (') and double prime (") have been omitted for convenience; k_t, κ_0, and $\bar{\rho}$ are respectively defined as

$$k_t = \hat{x}k_x + \hat{y}k_y; \quad k_t = \sqrt{k_t \cdot k_t}; \quad \kappa_0 = \sqrt{k_0^2 - k_t^2}, \quad (24)$$

and

$$\bar{\rho} = \hat{x}x + \hat{y}y; \quad r = \bar{\rho} + \hat{z}z; \quad k_0^2 = \omega^2\mu_0\epsilon_0. \quad (25)$$

In the above formulation, the variable k_t (i.e., $\hat{x}k_x + \hat{y}k_y$; $dk_t = dk_x\,dk_y$) are the continuous eigenvalues which span over the entire spectral domain ($-\infty < k_x < \infty$; and $-\infty < k_y < \infty$). The unknown spectral amplitudes $a'^{\gtrless}$ and $a'^{\gtrless}$ of (20) associated with the TM and TE modal fields respectively, are found from an application of the Lorentz reciprocity theorem to the pair of the fields $(E^{\gtrless}, H^{\gtrless})$ of (20) and the source free solenoidal vector wavefunctions $(e^{\gtrless}, h^{\gtrless})$ in the region V_0, bounded by planar surfaces $S_>$ and $S_<$, which are slightly above and below the surface S of Fig. 2, respectively, [13], [14][1]:

$$\iint_{S_> + S_<} ds \cdot (E^{\gtrless} \times h^{\lessgtr} - e^{\lessgtr} \times H^{\gtrless})$$

$$= \iiint_{V_0} dv\, e^{\lessgtr} \cdot J_0$$

$$= e^{\lessgtr} \cdot p_e. \quad (26)$$

The solenoidal vector wavefunctions $e^{\gtrless}$ and $h^{\gtrless}$ satisfy the orthogonality condition on the surface $S_<$ and $S_>$; namely,

$$\iint_{S_{\gtrless}} ds \cdot (e^{<}(\pm k_t) \times h^{>}(\mp k'_t))$$

$$= \iint_{S_{\gtrless}} ds \cdot (-e^{>}(\pm k_t) \times h^{<}(\mp k'_t))$$

$$= (\hat{s}_{\gtrless} \cdot \hat{z})\Omega\,\delta(k_t - k'_t), \quad (27)$$

where Ω can be Ω' or Ω'' for the TM and TE cases, respectively; thus,

$$\Omega' = k_t^2\eta'_0; \quad \Omega'' = \frac{k_t^2}{\eta''_0}, \quad (28)$$

[1]We apply the Lorentz reciprocity theorem to the volume V_0 here with the radiation condition implied as $\rho \to \infty$.

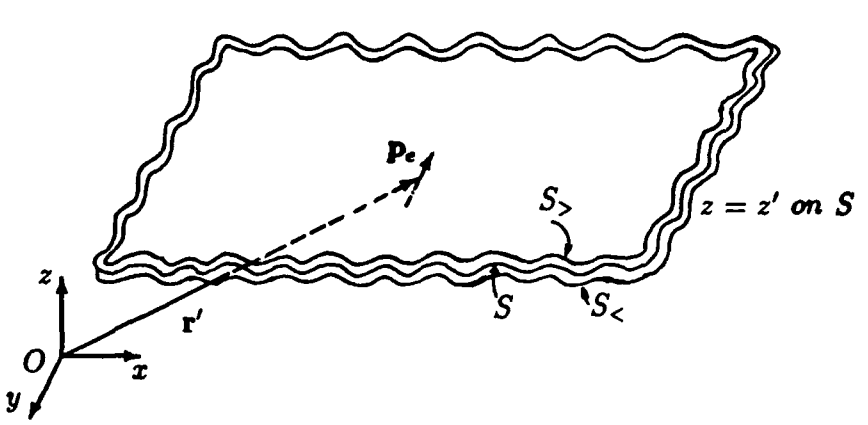


Fig. 2. Imaginary plane S, parallel to xy plane, passing through the source at $z = z'$ in the free space. Also plotted are planar surfaces $S_>$ and $S_<$ slightly above and below the source.

with the unit vector $\hat{s}_\gtrless$ directed along the outward normal to the surface $S_\gtrless = \pm z$; and η_0' and η_0'' are associated with the TM and TE wave impedances for region (0) and defined as

$$\eta_0' = \frac{\kappa_0}{\omega\epsilon_0}; \quad \eta_0'' = \frac{\omega\mu_0}{\kappa_0}. \tag{29}$$

Incorporating (20) and (27) into (26) yields

$$a^\gtrless = \frac{e^\gtrless(-k_t, r') \cdot p_e}{-2\Omega}; \quad r' = \bar{\rho}' + \hat{z}z'. \tag{30}$$

In deriving the orthogonality relationship of (27), use has been made of

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi e^{-j\xi(\xi - \xi')} = \delta(\xi - \xi'). \tag{31}$$

Therefore, from (30), (20), (16), (6), and (4), the z-propagation PWS representation of the free space dyadic Green's function can be identified as

$$\begin{aligned}\mathcal{G}^0(r, r') = {} & \frac{\mathcal{U}(z - z')}{-j\omega\mu_0}\int dk_t \left(\frac{e'^{>}(k_t, r)e'^{<}(-k_t, r')}{-2\Omega'}\right. \\ & \left. + \frac{e''^{>}(k_t, r)e''^{<}(-k_t, r')}{-2\Omega''}\right) \\ & + \frac{\mathcal{U}(z' - z)}{-j\omega\mu_0}\int dk_t \left(\frac{e'^{<}(k_t, r)e'^{>}(-k_t, r')}{-2\Omega'}\right. \\ & \left. + \frac{e''^{<}(k_t, r)e''^{>}(-k_t, r')}{-2\Omega''}\right) - \frac{\hat{z}\hat{z}}{k_0^2}\delta(r - r').\end{aligned} \tag{32}$$

The PWS for the fields (E_0 and H_0) due to p_e in free space, and hence for the corresponding free space dyadic Green's function $\mathcal{G}^0$ given above provides information on the general form of the PWS solution for the fields E_m and H_m for the multilayered case and therefore also on the dyadic Green's function for the multilayered media, which will be discussed in the following section.

IV. Construction of the Multilayered Dyadic Green's Function, $\mathcal{G}^{m,0}$

The electric dyadic Green's function for the multilayered media can be expressed as a sum of $\mathcal{G}^0$ in (32) and another contribution to account for the field scattered by the layered media. The scattered contribution can be expressed in terms of a PWS integral resembling that for $\mathcal{G}^0$. Let us consider an arbitrarily oriented point dipole source in a general multilayered media with constitutive parameters μ_m and ϵ_m, as shown in Fig. 1. The source is located in region (0) with constitutive parameters μ_0 and ϵ_0. In order to find the explicit value of the fields in each region, one can write the field quantities as the superposition of the four traveling waves (two oppositely traveling waves for each mode) with unknown coefficients and then solve for the unknown coefficients by enforcing the continuity of tangential electromagnetic fields quantities at each interface, [7], [8], [21]. However, we pursue another approach, which provides a useful physical interpretation for the dyadic Green's function. From (21) and (22), one can see that the continuity of the tangential field quantities at the interface m imply

$$\Pi'_{m-1} = \Pi'_m; \quad \frac{1}{\epsilon_{m-1}}\frac{\partial}{\partial z}\Pi'_{m-1} = \frac{1}{\epsilon_m}\frac{\partial}{\partial z}\Pi'_m, \tag{33}$$

and

$$\Pi''_{m-1} = \Pi''_m; \quad \frac{1}{\mu_{m-1}}\frac{\partial}{\partial z}\Pi''_{m-1} = \frac{1}{\mu_m}\frac{\partial}{\partial z}\Pi''_m. \tag{34}$$

These boundary conditions are analogous to the continuity of the current and voltage at each discontinuity of a piecewise uniform transmission line for which the characteristic impedance (and the wave number) is defined in each layer as, [6], [9]

$$\eta_m' = \frac{\kappa_m}{\omega\epsilon_m}; \quad \eta_m'' = \frac{\omega\mu_m}{\kappa_m}, \tag{35}$$

where prime (') and double prime (") are associated with TM_z and TE_z cases respectively, and $\kappa_m = \sqrt{k_m^2 - k_t^2}$, is the wave number in the $\hat{z}$-direction.

The field quantities in region m can be expressed as a superposition of known continuous solenoidal eigenfunctions that propagate in $\pm\hat{z}$-direction with the unknown spectral weights, a_m, [9]

$$\begin{aligned} \boldsymbol{E}_m^{\gtrless} &= \int d\boldsymbol{k}_t \, (a_m'^{\gtrless}(\boldsymbol{e}_m'^{\gtrless} + R_m'^{\gtrless}(0)\boldsymbol{e}_m'^{\lessgtr}) \\ &\quad + a_m''^{\gtrless}(\boldsymbol{e}_m''^{\gtrless} + R_m''^{\gtrless}(0)\boldsymbol{e}_m''^{\lessgtr})), \\ \boldsymbol{H}_m^{\gtrless} &= \int d\boldsymbol{k}_t \, (a_m'^{\gtrless}(\boldsymbol{h}_m'^{\gtrless} + R_m'^{\gtrless}(0)\boldsymbol{h}_m'^{\lessgtr}) \\ &\quad + a_m''^{\gtrless}(\boldsymbol{h}_m''^{\gtrless} + R_m''^{\gtrless}(0)\boldsymbol{h}_m''^{\lessgtr})), \end{aligned} \tag{36}$$

where $R_m^{\gtrless}(0) = R_m^{\gtrless} e^{\mp j2\kappa_m z_m}$, $R_m'^{\gtrless}$ and $R_m''^{\gtrless}$ are the TM, and TE effective reflection coefficients at the interfaces $(m, m+1)$ and $(m, m-1)$ for $(>)$ and $(<)$, respectively [7], [8], [18], [19], [21]. As discussed in the Appendix, the effective reflection coefficient $R_m^{\gtrless}$ for region m, is a function of reflection coefficients of all successive layers, (i.e., $m \pm 1$, $m \pm 2$, $m \pm 3$, $\cdots$,; $(\substack{+ \text{ for } > \\ - \text{ for } <})$) of the multilayered media, (in particular see (A13) and (A14)). Also the modal coefficients $a_m^{\gtrless}$ of region m, and $a_n^{\gtrless}$ of region n on either side of the source are related via the effective transmission coefficient, $T_{m,n}^{\gtrless}$, as is evident from the piece-wise transmission line theory discussed in the Appendix. In view of (A11), (A12), (A16), (21) and (22), one will have[2]

$$a_m^{\gtrless} = e^{\mp j(\kappa_n z_n - \kappa_m z_m)} \Theta_{m,n}^{\gtrless} a_n^{\gtrless};$$

$$\text{where} \begin{cases} \Theta_{m,n}'^{\gtrless} = \dfrac{\eta_{n'}}{\eta_{m'}} T_{m,n}'^{\gtrless}, \\ \qquad \text{for TM } (') \text{ case} \\ \Theta_{m,n}''^{\gtrless} = T_{m,n}''^{\gtrless}, \\ \qquad \text{for TE } (") \text{ case.} \end{cases} \tag{37}$$

Hence, one only needs to find the modal coefficients $a_0'^{\gtrless}$, and $a_0''^{\gtrless}$ in region 0, in order to completely specify the fields in all regions. Specifying (36) for region 0 (i.e., $m = 0$) and invoking the Lorentz reciprocity theorem to the pair of fields $(\boldsymbol{E}_0^{\gtrless}, \boldsymbol{H}_0^{\gtrless})$ and a set of source free test fields $(\mathcal{E}_0^{\lessgtr}, \mathcal{H}_0^{\lessgtr})$ in the region V_0, bounded by the planar surfaces of $S_<$ and $S_>$, slightly below and above the source respectively as shown in Fig. (1), we get

$$\begin{aligned} &\iint_{S_> + S_<} d\boldsymbol{s} \cdot (\boldsymbol{E}_0^{\gtrless} \times \mathcal{H}_0^{\lessgtr} - \mathcal{E}_0^{\lessgtr} \times \boldsymbol{H}_0^{\gtrless}) \\ &\quad = \iiint_{V_0} dv \, \mathcal{E}_0^{\lessgtr} \cdot \boldsymbol{J}_0 \\ &\quad = \mathcal{E}_0^{\lessgtr} \cdot \boldsymbol{p}_e. \end{aligned} \tag{38}$$

[2]Note that $T_{m,n}^{\gtrless}$ is the ratio of the incident tangential electrical fields of regions m and n; namely,

$$a_m^{\gtrless}(\hat{z} \times \boldsymbol{e}_m^{\gtrless}) = e^{\mp j(\kappa_n z_n - \kappa_m z_m)} T_{m,n}^{\gtrless} a_n^{\gtrless}(\hat{z} \times \boldsymbol{e}_n^{\gtrless}).$$

where $(\mathcal{E}_0^{\lessgtr}, \mathcal{H}_0^{\lessgtr})$ are given by

$$\mathcal{E}_0^{\lessgtr} = \boldsymbol{e}_0^{\lessgtr} + R_0^{\lessgtr}(0)\boldsymbol{e}_0^{\gtrless}; \quad \mathcal{H}_0^{\lessgtr} = \boldsymbol{h}_0^{\lessgtr} + R_0^{\lessgtr}(0)\boldsymbol{h}_0^{\gtrless},$$

and

$$R_0^{\lessgtr}(0) = R_0^{\lessgtr}(z_0^{\mp}) e^{\pm j2\kappa_0 z_0^{\pm}}. \tag{39}$$

The solenoidal vector functions $(\mathcal{E}_0^{\lessgtr}, \mathcal{H}_0^{\lessgtr})$ satisfy the orthogonality relationship on the planar surface of $S_{\gtrless}$ as is evident from (27),

$$\begin{aligned} &\iint_{S_{\gtrless}} d\boldsymbol{s} \cdot (\mathcal{E}_0^{<}(\pm\boldsymbol{k}_t) \times \mathcal{H}_0^{>}(\mp\boldsymbol{k}_t') - \mathcal{E}_0^{>}(\mp\boldsymbol{k}_t) \\ &\qquad \times \mathcal{H}_0^{<}(\pm\boldsymbol{k}_t')) = (\hat{s}_{\gtrless} \cdot \hat{z}) 2\Lambda \, \delta(\boldsymbol{k}_t - \boldsymbol{k}_t'), \end{aligned} \tag{40}$$

with

$$\Lambda = \Omega(1 - R_0^{>}(z_0^{+}) R_0^{<}(z_0^{-}) e^{-j2\kappa_0 d_0}), \tag{41}$$

where $d_0 = z_0^{+} - z_0^{-}$ is the thickness of the slab 0; z_0^{+} and z_0^{-} are the values of z at the interfaces of region (0) and they are specified in Fig. 1, and Ω is given in (28). Incorporating (36) and (40) into (38) yields

$$a_0^{\gtrless} = \frac{\mathcal{E}_0^{\lessgtr}(-\boldsymbol{k}_t, \boldsymbol{r}') \cdot \boldsymbol{p}_e}{-2\Lambda}. \tag{42}$$

The prime (') and double prime (") have been omitted for convenience in (37)–(42). Hence, the electric dyadic Green's function for the multilayered media, $\mathcal{G}^{m,0}$, can be written via (42), (39), (37), (36), (16) and (4) as

$$\begin{aligned} \mathcal{G}^{m,0}(\boldsymbol{r}, \boldsymbol{r}') &= \frac{\mathcal{U}(z - z')}{-j\omega\mu_0} \int d\boldsymbol{k}_t \\ &\quad \cdot \left(\Theta_{m,0}'^{>} \frac{\mathcal{E}_m'^{>}(\boldsymbol{k}_t, \boldsymbol{r}) \mathcal{E}_0'^{<}(-\boldsymbol{k}_t, \boldsymbol{r}')}{-2\Lambda'} \right. \\ &\quad \left. + \Theta_{m,0}''^{>} \frac{\mathcal{E}_m''^{>}(\boldsymbol{k}_t, \boldsymbol{r}) \mathcal{E}_0''^{<}(-\boldsymbol{k}_t, \boldsymbol{r}')}{-2\Lambda''} \right) \\ &\quad + \frac{\mathcal{U}(z' - z)}{-j\omega\mu_0} \int d\boldsymbol{k}_t \\ &\quad \cdot \left(\Theta_{m,0}'^{<} \frac{\mathcal{E}_m'^{<}(\boldsymbol{k}_t, \boldsymbol{r}) \mathcal{E}_0'^{>}(-\boldsymbol{k}_t, \boldsymbol{r}')}{-2\Lambda'} \right. \\ &\quad \left. + \Theta_{m,0}''^{<} \frac{\mathcal{E}_m''^{<}(\boldsymbol{k}_t, \boldsymbol{r}) \mathcal{E}_0''^{>}(-\boldsymbol{k}_t, \boldsymbol{r}')}{-2\Lambda''} \right) \\ &\quad - \frac{\hat{z}\hat{z}}{k_0^2} \delta(\boldsymbol{r} - \boldsymbol{r}'). \end{aligned} \tag{43}$$

where $(\mathcal{E}_m^{\lessgtr}, \mathcal{H}_m^{\lessgtr})$ are given by $(\mathcal{E}_m^{\lessgtr} = \boldsymbol{e}_m^{\lessgtr} + R_m^{\lessgtr}(0)\boldsymbol{e}_m^{\gtrless}$; $\mathcal{H}_m^{\lessgtr} = \boldsymbol{h}_m^{\lessgtr} + R_m^{\lessgtr}(0)\boldsymbol{h}_m^{\gtrless})$. The above expression for $\mathcal{G}^{m,0}$ can be written more explicitly using (39), (37) and (21)–(23) as

$$\begin{aligned} \mathcal{G}^{m,0} &= -\frac{\hat{z}\hat{z}}{k_0^2} \delta(\boldsymbol{r} - \boldsymbol{r}') + \mathcal{U}(z - z') \frac{-j}{4\pi^2} \int d\boldsymbol{k}_t \\ &\quad \cdot e^{j\kappa_m z_m} e^{-j\kappa_0 z_0^{+}} \frac{1}{2\kappa_0} [\Delta'' T_{m,0}''^{>} \hat{n}'' \hat{n}'' \\ &\quad \cdot (e^{-j\boldsymbol{k}_m^{>} \cdot \boldsymbol{r}} + R_m''^{>} e^{-j2\kappa_m z_m} e^{-j\boldsymbol{k}_m^{<} \cdot \boldsymbol{r}}) \end{aligned}$$

$$\cdot (e^{+jk_0^{>}\cdot r'} + R_0''^{<} e^{+j2\kappa_0 z_0^-} e^{+jk_0^{<}\cdot r'})$$

$$+ \Delta' T'^{>}_{m,0} \frac{k_m \kappa_0}{k_0 \kappa_m} (\hat{n}_m'^{>} e^{-jk_m^{>}\cdot r}$$

$$+ R_m'^{>} e^{-j2\kappa_m z_m} \hat{n}_m'^{<} e^{-jk_m^{<}\cdot r})(\hat{n}_0'^{>} e^{+jk_0^{>}\cdot r'}$$

$$+ R_0'^{<} e^{+j2\kappa_0 z_0^-} \hat{n}_0'^{<} e^{+jk_0^{<}\cdot r'})]$$

$$+ \mathcal{U}(z' - z) \cdot \frac{-j}{4\pi^2} \int dk_t$$

$$\cdot e^{-j\kappa_m z_m} e^{j\kappa_0 z_0^-} \frac{1}{2\kappa_0} [\Delta'' T''^{<}_{m,0} \hat{n}'' \hat{n}''$$

$$\cdot (e^{-jk_m^{<}\cdot r} + R_m''^{<} e^{+j2\kappa_m z_m} e^{-jk_m^{>}\cdot r})$$

$$\cdot (e^{+jk_0^{<}\cdot r'} + R_0''^{>} e^{-j2\kappa_0 z_0^+} e^{+jk_0^{>}\cdot r'})$$

$$+ \Delta' T'^{<}_{m,0} \frac{k_m \kappa_0}{k_0 \kappa_m} (\hat{n}_m'^{<} e^{-jk_m^{<}\cdot r}$$

$$+ R_m'^{<} e^{+j2\kappa_m z_m} \hat{n}_m'^{>} e^{-jk_m^{>}\cdot r})$$

$$\cdot (\hat{n}_0'^{<} e^{+jk_0^{<}\cdot r'} + R_0'^{>} e^{-j2\kappa_0 z_0^+} \hat{n}_0'^{>} e^{+jk_0^{>}\cdot r'})], \quad (44)$$

where Δ is defined as (Ω/Λ), z_0^+, z_0^-, z_m are specified in Fig. 1, and $k_i^{\gtrless}$ is given by

$$k_i^{\gtrless} = \hat{x}k_x + \hat{y}k_y \pm \hat{z}\kappa_i; \quad |k_i^{\gtrless}| = k_i = \omega\sqrt{\mu_i \epsilon_i}. \quad (45)$$

and, unit vectors $\hat{n}''$, and $\hat{n}_i'$ are defined as

$$\hat{n}'' = \frac{\hat{x}k_y - \hat{y}k_x}{k_t}; \quad \hat{n}_i'^{\gtrless} = \frac{(-\hat{x}k_x - \hat{y}k_y)\kappa_i \pm \hat{z}k_t^2}{k_i k_t}. \quad (46)$$

Also note that $R_0^{>} = R^{>}(z_0^+)$, $R_0^{<} = R(z_0^-)$, and $R_m^{\gtrless} = R^{\gtrless}(z_m)$ for $m \gtrless 0$. The physical interpretation of the parameters defined here will be discussed in the following section.

V. Physical Interpretation of the Dyadic Green's Function for a Multilayered Media

In this section we will try to give some physical insight to the dyadic Green's function of the multilayered media derived in the preceding section.

The double prime, ($''$), denotes plane waves in the PWS representation for which the electric field is normal to the plane of incidence, (i.e., the plane defined by the propagation vector, k, and the direction normal $\hat{z}$); thus, the polarization of electric field vector, $\hat{n}''$, is given by

$$\hat{n}'' = \frac{-\hat{z} \times k}{|\hat{z} \times k|} = \frac{\hat{x}k_y - \hat{y}k_x}{k_t}. \quad (47)$$

Likewise the prime, ($'$), denotes plane waves with the electric field in the plane of incidence (with the magnetic field normal to the plane of incidence). In this case, the polarization of the electric field vector, $\hat{n}'$, is given by

$$\hat{n}'^{\gtrless} = \pm \frac{\hat{n}'' \times k^{\gtrless}}{|\hat{n}'' \times k|} = \frac{(-\hat{x}k_x - \hat{y}k_y)\kappa \pm \hat{z}k_t^2}{kk_t} \quad (48)$$

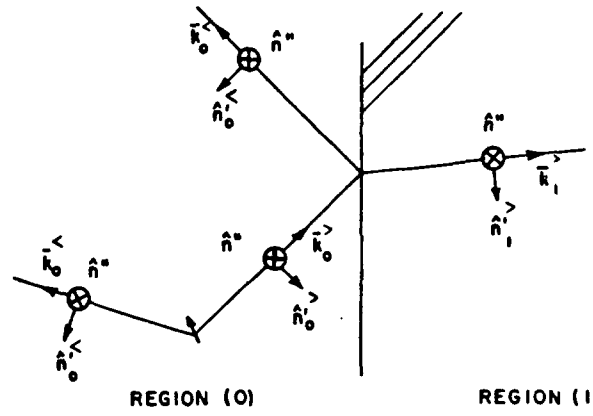


Fig. 3. Directions of $k_m^{\gtrless}$, $n_m'^{\gtrless}$ and n_m'' on either side of interface (0); ($m = 0, 1$).

where $\hat{n}'^{\gtrless}$ is for the wave traveling in the $\pm\hat{z}$ directions as shown in Fig. 3.

The dyadic Green's function evaluated in the region (m) consists of the spectrum of two types of plane waves excited by the source at $z = z'$ in region (0); these are the direct (incident) plus reflected waves. The total "effective" incident wave at $z = z_0^+$ is given by (see also (A11) of the Appendix),

$$e^{-j\kappa_0 z_0^+} \Delta (e^{jk_0^{>}\cdot r'} + R_0^{<} e^{j2\kappa_0 z_0^-} e^{jk_0^{<}\cdot r'});$$

$$\Delta = \frac{1}{1 - R_0^{>} R_0^{<} e^{-j2\kappa_0 d_0}}, \quad (49)$$

and Δ is the sum of the geometric series,

$$\Delta = 1 + \alpha + \alpha^2 + \alpha^3 + \cdots;$$

$$\alpha = R_0^{>}(z_0^+) R_0^{<}(z_0^-) e^{-j2\kappa_0 d_0}. \quad (50)$$

Physically Δ in (50) is the total sum of the plane waves traveling in $+\hat{z}$ or $-\hat{z}$ directions which result from the infinite number of bounces at the interfaces of slab (0), therefore it can be viewed as the "effective" incident wave at $z = z_0^+$, as is shown geometrically in Fig. 4.

The total incident wave at $z = z_0^+$ is transmitted through the slabs (0 to m), by the effective transmission coefficients, $T^{>}_{m,0}$ (see (A12) of the Appendix),

$$T^{>}_{m,0} = (T_0 e^{-jk_1 d_1})(T_2 e^{-jk_2 d_2}) \cdots (T_{m-1} e^{-jk_m d_m}), \quad (51)$$

where d_i is thickness of the slab (i), (for $i = 0$ to m). At the slab m, the total field will be the superposition of the effective incident field plus the effective reflected field from the boundary at $z = z_m$ as shown in Fig. 5.

Note that the ratio of $k_m\kappa_0/k_0\kappa_m$ in the TM ($'$) part of (44) is simply the ratio of the cosine of the angles that k_0 and k_m make with the normal of the interface which is depicted in Fig. 6, and results from the continuity of the tangential TM electric fields at the each interface.

Although the limits of the spectral integral extend from $-\infty$ to ∞, the reflection and transmission coefficients, Γ_m and τ_m, for each interface and hence, the effective reflection and transmission coefficients, R_m and T_m, have an asymptotic limit for large value of k_t. Figs. 7–10 show the

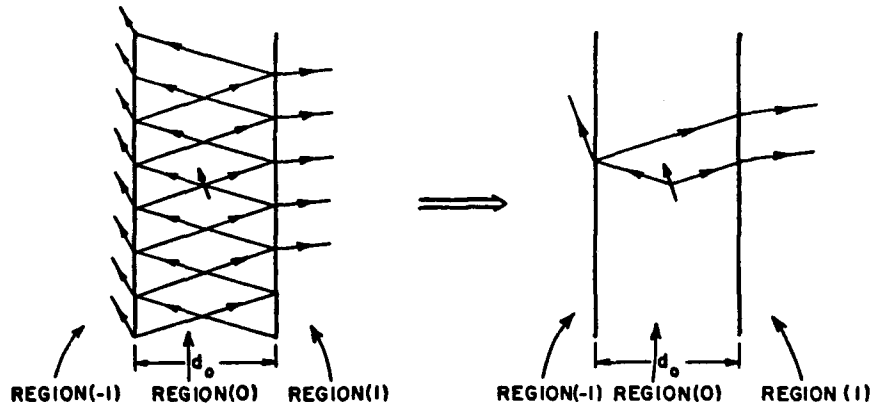


Fig. 4. Plane waves bouncing back and forth at the interfaces of the slab (0) and its equivalent representation.

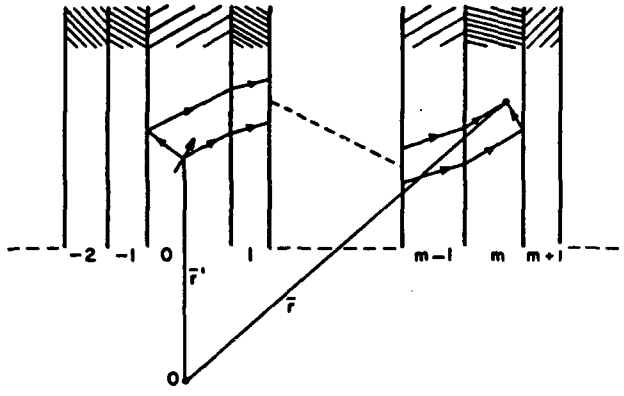


Fig. 5. Physical interpretation of incident and reflected waves in the slab m due to the point current dipole source in the slab (0).

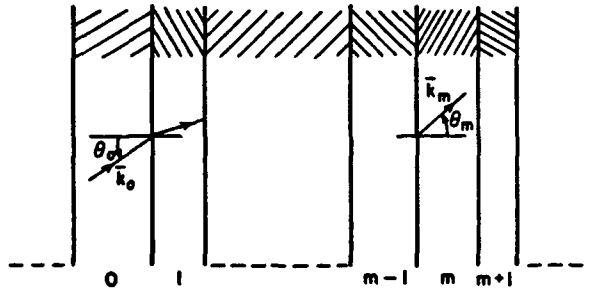


Fig. 6. Direction cosines that k_0 and k_m make with the normal $\hat{z}$, these result from the continuity of the tangential TM electric field at each interface; $\theta_0 = \cos^{-1}(\kappa_0/k_0)$, $\theta_m = \cos^{-1}(\kappa_m/k_m)$.

real and imaginary parts of effective reflection and transmission coefficients for one, two and three layer geometries as a function of normalized k_t, (with respect to the free space wave number, k_0), for TM and TE cases, respectively. It is evident that the values of these coefficients approach certain limiting constants for large values of k_t. It can be seen from Figs. 8 and 10 that the values of effective transmission coefficients approach zero for k_t larger than 3; physically this implies that no evanescent wave with a large transverse wave number k_t can penetrate through the layers. One can of course predict these phenomena by taking the limits of the reflection and transmission coefficients of (A12)–(A15) as k_t goes to infinity. That is,

$$R_m^{\gtrless} \approx \Gamma^{\gtrless}_{m|k_t \gg k_m};$$

$$\lim_{k_t \to \infty} R_m^{\gtrless} = \lim_{k_t \to \infty} \Gamma_m^{\gtrless} \to \begin{cases} \dfrac{\epsilon_m - \epsilon_{m\pm1}}{\epsilon_m + \epsilon_{m\pm1}} & \text{for TM } (') \\ \dfrac{\mu_{m\pm1} - \mu_m}{\mu_{m\pm1} + \mu_m} & \text{for TE } ('') \end{cases} \tag{52}$$

and

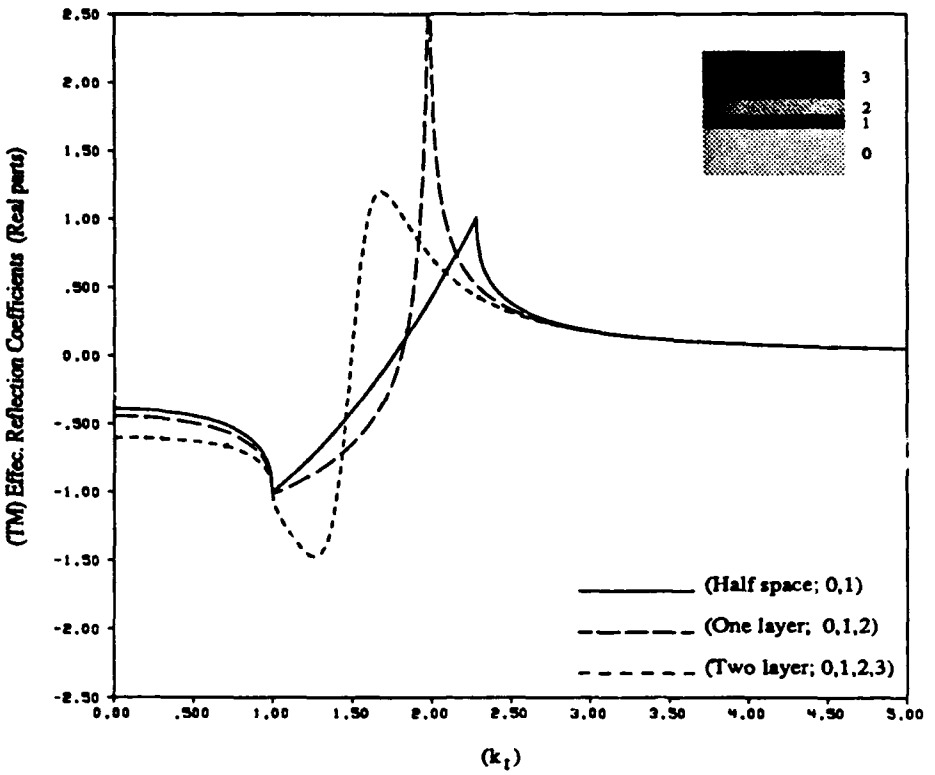
$$T_m^{\gtrless} \approx \tau^{\gtrless}_{m|k_t \gg k_m};$$

$$\lim_{k_t \to \infty} T_m^{\gtrless} = \lim_{k_t \to \infty} \tau_m^{\gtrless} \to \begin{cases} \dfrac{2\epsilon_m}{\epsilon_m + \epsilon_{m\pm1}} & \text{for TM } (') \\ \dfrac{2\mu_{m\pm1}}{\mu_{m\pm1} + \mu_m} & \text{for TE } (''). \end{cases} \tag{53}$$

As is evident from (51) for any multilayered media with a nonzero thickness, we will have

$$\lim_{k_t \to \infty} T^{\gtrless}_{m,0} \longrightarrow 0. \tag{54}$$

The numerical implication of this phenomena is that for large values of k_t, the effective reflection and transmission coefficients, R_m and T_m, can be replaced by their associated half-space reflection and transmission coefficients, Γ_m and τ_m. Also it is evident from Figs. 7 and 8 that for a set of constitutive parameters and layer thicknesses, there exist some value of k_t for which the denominators of effective reflection and transmission coefficients go to zero and consequently these coefficients become singular. These values of k_t correspond to the surface wave modes, and the associated residues are proportional to the fields of these modes where are launched by the impressed source [6], [20]. Also, the sharp variation of these effective reflection and transmission coefficients, at the various points



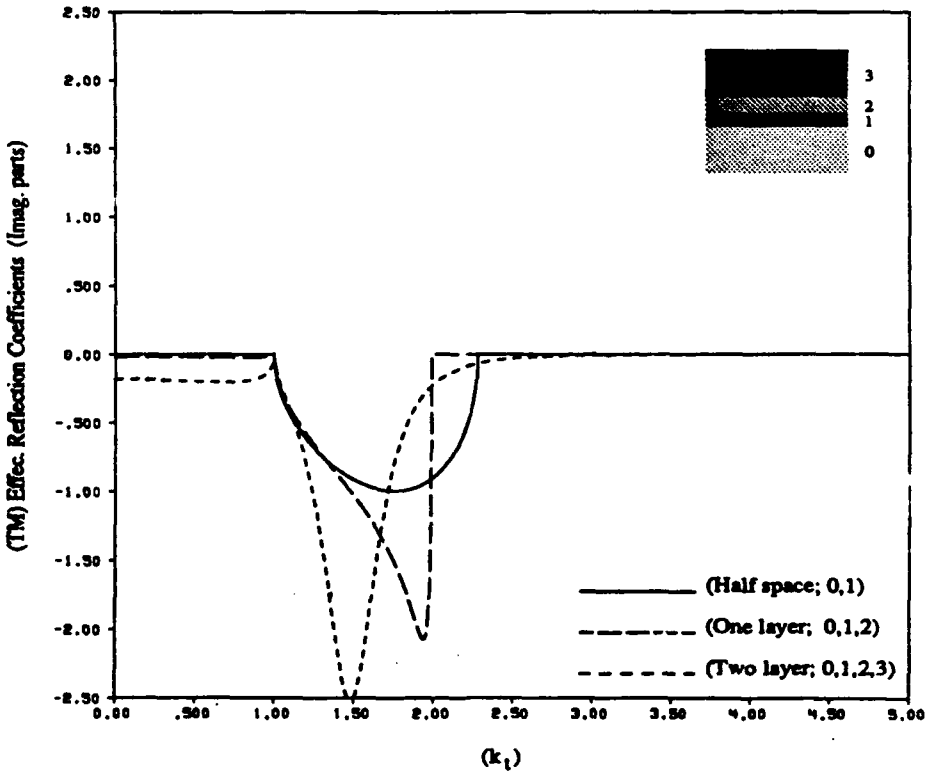
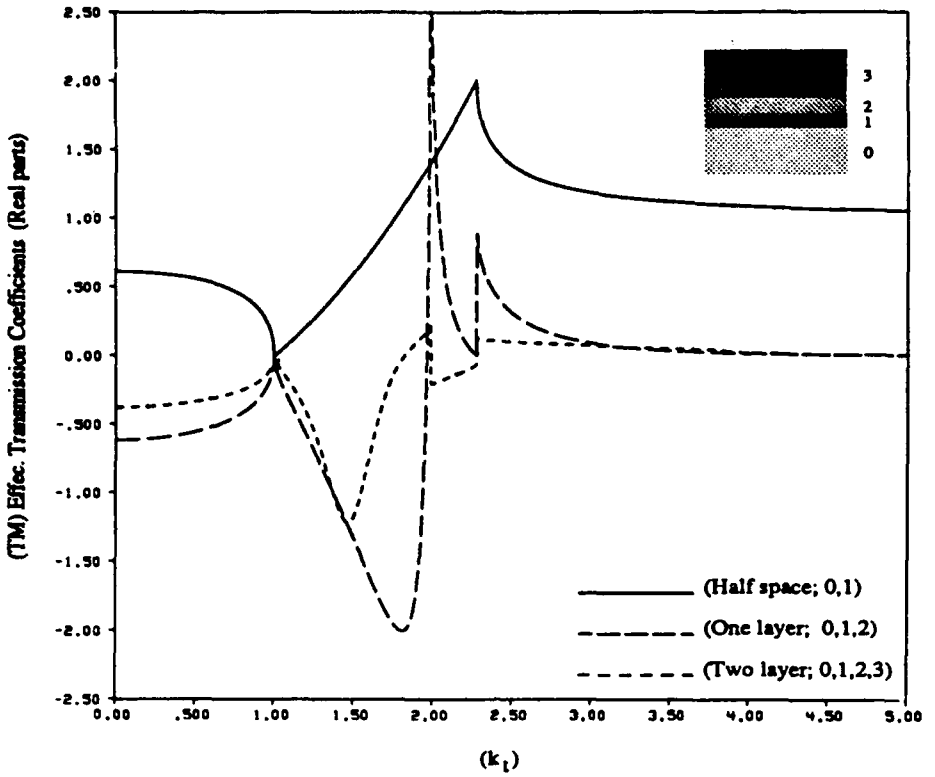


Fig. 7. Real and imaginary parts of effective TM reflection coefficients as a function of normalized k_t (with respect to k_0) for a half-space, as well as for one-layer and two-layer media on a half-space. The relative constitutive parameters and layer thicknesses are: ($\mu_{0,r} = 1.0$, $\epsilon_{0,r} = 1.0$), ($\mu_{1,r} = 1.2$, $\epsilon_{1,r} = 3.25$), ($\mu_{2,r} = 1.3$, $\epsilon_{2,r} = 10.2$), ($\mu_{3,r} = 1.6$, $\epsilon_{3,r} = 2.2$), ($d_1/\lambda_0 = 0.1$), and ($d_2/\lambda_0 = 0.1$).



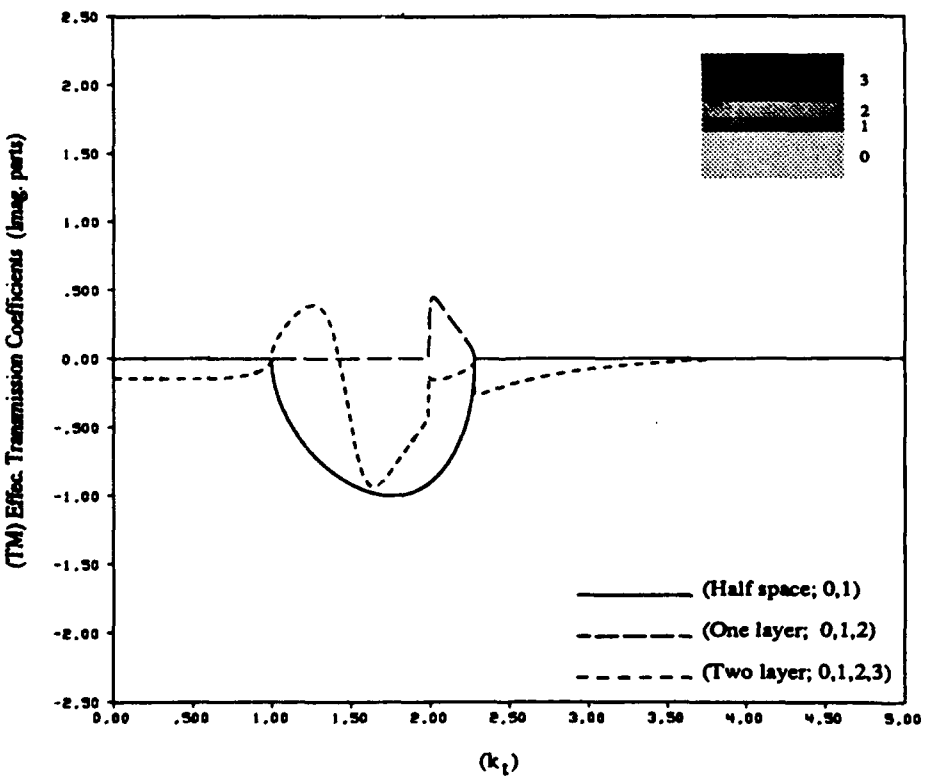
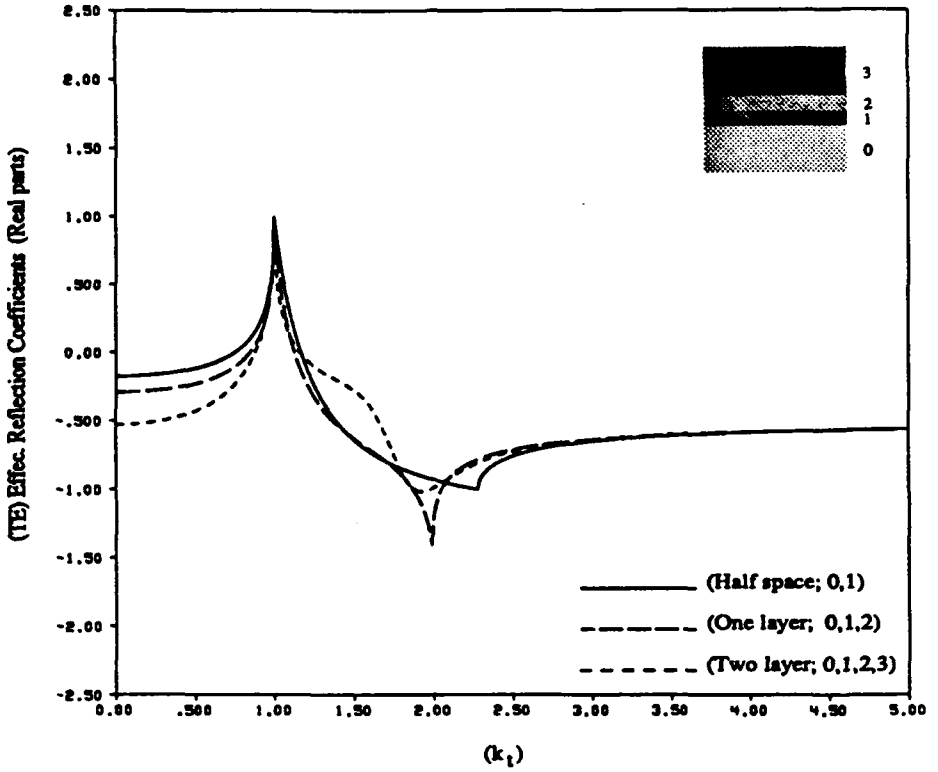


Fig. 8. Real and imaginary parts of effective TM transmission coefficients as a function of normalized k_t (with respect to k_0) for a half-space, as well as for one-layer and two-layer media on a half-space. The relative constitutive parameters and layer thicknesses are: ($\mu_{0,r} = 1.0$, $\epsilon_{0,r} = 1.0$), ($\mu_{1,r} = 1.2$, $\epsilon_{1,r} = 3.25$), ($\mu_{2,r} = 1.3$, $\epsilon_{2,r} = 10.2$), ($\mu_{3,r} = 1.6$, $\epsilon_{3,r} = 2.2$), ($d_1/\lambda_0 = 0.1$), and ($d_2/\lambda_0 = 0.1$).



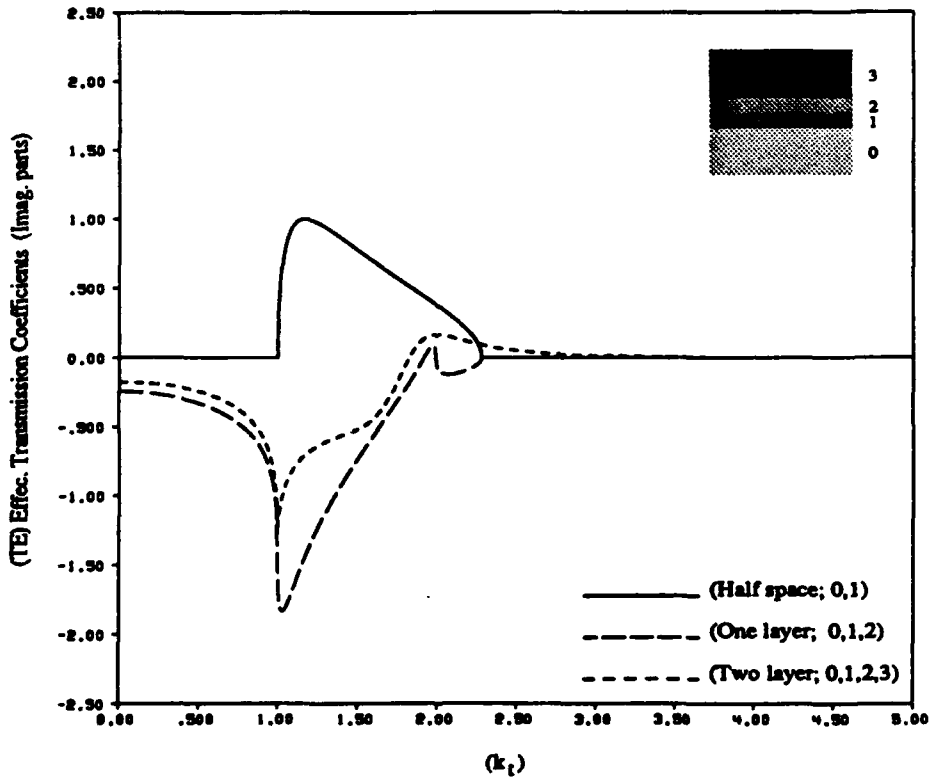
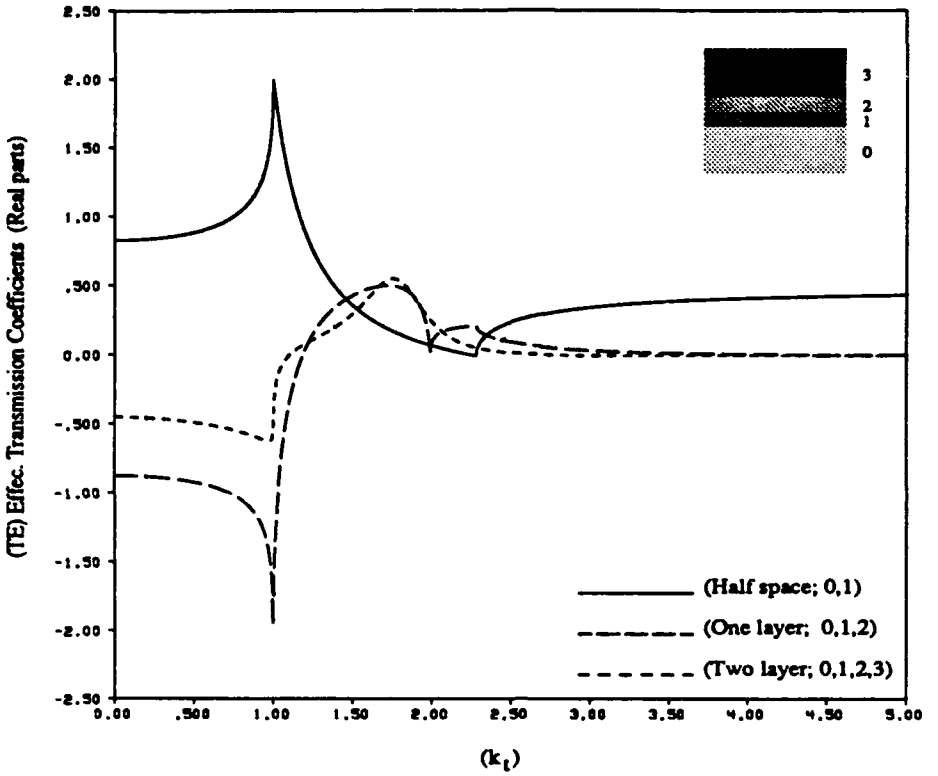


Fig. 9. Real and imaginary parts of effective TE reflection coefficients as a function of normalized k_t (with respect to k_0) for a half-space, as well as for one-layer and two-layer media on a half-space. The relative constitutive parameters and layer thicknesses are: ($\mu_{0,r} = 1.0$, $\epsilon_{0,r} = 1.0$), ($\mu_{1,r} = 1.2$, $\epsilon_{1,r} = 3.25$), ($\mu_{2,r} = 1.3$, $\epsilon_{2,r} = 10.2$), ($\mu_{3,r} = 1.6$, $\epsilon_{3,r} = 2.2$), ($d_1/\lambda_0 = 0.1$), and ($d_2/\lambda_0 = 0.1$).



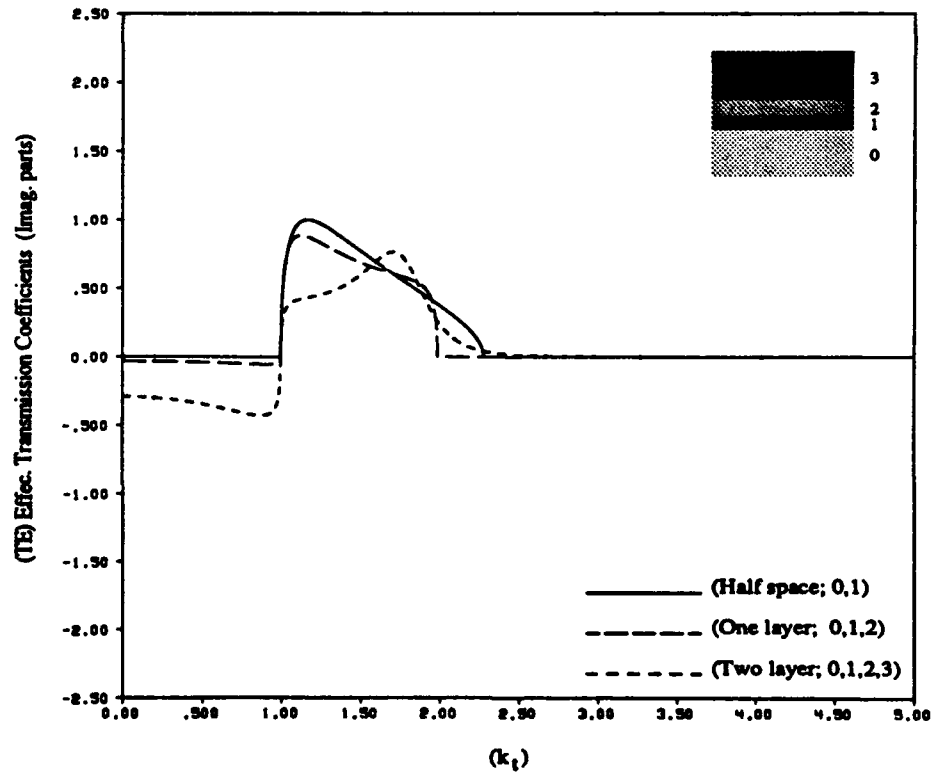


Fig. 10. Real and imaginary parts of effective TE transmission coefficients as a function of normalized k, (with respect to k_0) for a half-space, as well as for one-layer and two-layer media on a half-space. The relative constitutive parameters and layer thicknesses are: ($\mu_{0,r} = 1.0$, $\epsilon_{0,r} = 1.0$), ($\mu_{1,r} = 1.2$, $\epsilon_{1,r} = 3.25$), ($\mu_{2,r} = 1.3$, $\epsilon_{2,r} = 10.2$), ($\mu_{3,r} = 1.6$, $\epsilon_{3,r} = 2.2$), ($d_1/\lambda_0 = 0.1$), and ($d_2/\lambda_0 = 0.1$).

in these Figures result from an abrupt phase change of the associated half-space reflection and transmission coefficients at the vicinity of $k_t = k_m$, for (m = 0, 1, 2, or 3).

VI. Conclusion

A relatively simple and systematic approach is taken to drive the dyadic Green's function for a multilayered dielectric/magnetic media via the two ($\hat{z}$)-directed solenoidal eigenfunctions, and the utilization of the Lorentz reciprocity theorem such that it provides a useful physical interpretation. It is shown that the Green's dyadic can be written in terms of the spectrum of plane waves (TE and TM) which resemble the response of a source excited multiconnected piece-wise uniform transmission line. The concept of effective reflection and transmission coefficients is discussed, and the physical interpretation of the individual terms along with the limiting behavior of some of these terms is given.

Appendix
Piecewise Uniform Transmission Line Theory

In this Appendix we briefly review the piecewise uniform transmission line theory. As explained earlier, the z and z' functional dependence of the field quantities excited by a electric point dipole current source in a general multilayered media is analogous to the problem of source excitation of a piecewise uniform transmission line. The voltage and current on a source free uniform transmission line with wave number κ_m and characteristic impedance η_m can be expressed as

$$V_m(z) = V_{inc,m}(z_0)(e^{-j\kappa_m(z-z_0)} + R_m(z_0)e^{j\kappa_m(z-z_0)}),$$

$$I_m(z) = \frac{V_{inc,m}(z_0)}{\eta_m}(e^{-j\kappa_m(z-z_0)} - R_m(z_0)e^{j\kappa_m(z-z_0)}), \quad \text{(A1)}$$

where $V_{inc,m}(z_0)$ and $R_m(z_0)$ are the incident voltage and reflection coefficient respectively at point $z = z_0$. The reflection coefficient, $R_m(z)$, and the impedance, $Z_m(z)$, at a point z are related by

$$R_m(z) = \frac{Z_m(z) - \eta_m}{Z_m(z) + \eta_m}; \quad Z_m(z) = \frac{V_m(z)}{I_m(z)}. \quad \text{(A2)}$$

It is desired to derive some expressions for a piecewise uniform transmission line that relate the voltages and currents at a pair of points on the line which are located in different sections. Let us first consider a simple configuration shown in Fig. 11 which consists of two semi-infinite transmission lines corresponding to regions (n − 1) and (n + 1), connected with a finite line, $d_n = z_n - z_{n-1}$, corresponding to region (n). For a known incident voltage in region (n − 1), it is of interest to find voltages and currents in different sections of the transmission line. For doing so, one needs to find the incident voltage and reach region. The reflection coefficient at z_{n-1} in region (n − 1) can be written as

$$R_{n-1} = \frac{Z(z_{n-1}) - \eta_{n-1}}{Z(z_{n-1}) + \eta_{n-1}}, \quad \text{(A3)}$$

where

$$Z(z_{n-1}) = \eta_n \frac{1 + R_n e^{-j2\kappa_n d_n}}{1 - R_n e^{-j2\kappa_n d_n}}; \quad R_n = \frac{\eta_{n+1} - \eta_n}{\eta_{n+1} + \eta_n}. \quad \text{(A4)}$$

After incorporating (A4) into (A3), the reflection coefficient R_{n-1} can be expressed as

$$R_{n-1} = \frac{\Gamma_{n-1} + R_n e^{-j2\kappa_n d_n}}{1 + \Gamma_{n-1} R_n e^{-j2\kappa_n d_n}}; \quad \Gamma_{n-1} = \frac{\eta_n - \eta_{n-1}}{\eta_n + \eta_{n-1}}. \quad \text{(A5)}$$

Expression R_{n-1} in (A5) is called "effective" reflection coefficient for region (n − 1). It is a coefficient that relates all interactions from the presence of other regions to the incident voltage in region (n − 1). One can also relate the incident voltages of regions (n − 1) and (n) in the following form:

$$V_{n-1}(z_{n-1}) = V_{inc,n-1}(z_{n-1})(1 + R_{n-1}(z_{n-1}))$$
$$= V_{inc,n}(z_{n-1})(1 + R_n e^{-j2\kappa_n d_n})), \quad \text{(A6)}$$

hence; $V_{inc,n}(z_{n-1})$ can be expressed in terms of $V_{inc,n-1}(z_{n-1})$ as

$$V_{inc,n}(z_{n-1}) = T_{n-1}(z_{n-1})V_{inc,n-1}(z_{n-1}), \quad \text{(A7)}$$

where

$$T_{n-1} = \frac{1 + R_{n-1}}{1 + R_n e^{-j2\kappa_n d_n}}. \quad \text{(A8)}$$

After substituting (A5) for R_{n-1} in (A8), T_{n-1} can be expressed as

$$T_{n-1} = \frac{\tau_{n-1}}{1 + \Gamma_{n-1} R_n e^{-j2\kappa_n d_n}}; \quad \tau_{n-1} = 1 + \Gamma_{n-1}. \quad \text{(A9)}$$

T_{n-1} in (A9) is called "effective" transmission coefficient. It is a coefficient that relates the incident wave of region (n − 1) to the incident wave of region (n). Therefore, the voltage at a point z in region (n) can be expressed in terms of the incident voltage at point z_0 in region (n − 1) by incorporating (A7) into (A1); hence,

$$V_n(z) = V_{inc,n-1}(z_0)e^{-j\kappa_{n-1}(z_{n-1}-z_0)}T_{n-1}e^{-j\kappa_n d_n}$$
$$\cdot (e^{-j\kappa_n(z-z_n)} + R_n e^{+j\kappa_n(z-z_n)}). \quad \text{(A10)}$$

The incident voltage in region (n + 1) can likewise be found in terms of the voltage in region (n).

This formulation can be generalized to the total of (N + 2) number of finite length transmission lines, (0^{th} and (N + 1)th regions are semi-infinite), with the characteristic impedance and wavenumber of η_m and κ_m, respectively for ($0 \le m \le N + 1$), as shown in Fig. 12. The

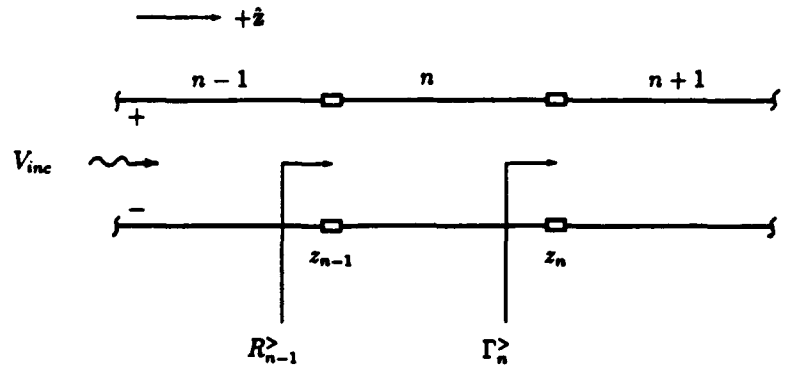


Fig. 11. Two infinite transmission lines connected with a finite length transmission line at $z = z_{n-1}$, and $z = z_n$.

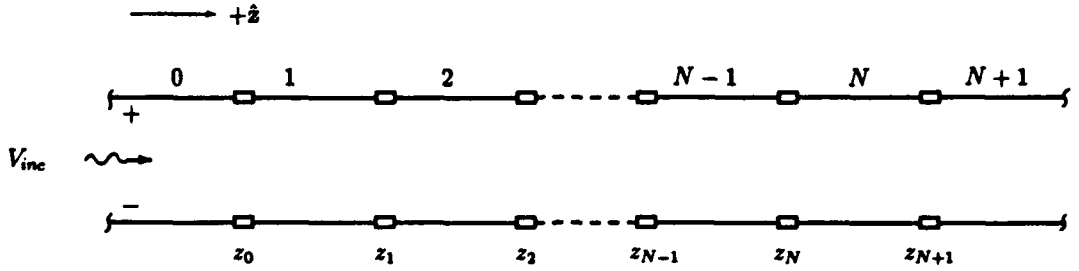


Fig. 12. General piecewise uniform transmission line; incident waves travel in $+\hat{z}$-direction.

voltage and current at point z in region (m), as a function of the incident voltage at point $z = 0$ in region (0) can be written as

$$V_m(z) = V^{>}_{inc,0}(0)e^{-jk_0z_0}\Upsilon^{>}_{m,0}(e^{-jk_m(z-z_m)} + R^{>}_m e^{+jk_m(z-z_m)}),$$

$$I_m(z) = \frac{V^{>}_{inc,0}(0)}{\eta_m} e^{-jk_0z_0}\Upsilon^{>}_{m,0}(e^{-jk_m(z-z_m)} - R^{>}_m e^{+jk_m(z-z_m)}), \quad \text{(A11)}$$

where, $\Upsilon^{>}_{m,0}$ and $R^{>}_m$ are respectively defined as

$$\Upsilon^{>}_{m,0} = \prod_{i=0}^{m-1} T^{>}_i(z_i)e^{-jk_{i+1}d_{i+1}};$$

$$T^{>}_i(z_i) = \frac{\tau^{>}_i}{1 + \Gamma^{>}_i R^{>}_{i+1} e^{-j2k_{i+1}d_{i+1}}}, \quad \text{(A12)}$$

and

$$R^{>}_m(z_m) = \frac{\Gamma^{>}_m + R^{>}_{m+1}e^{-j2k_{m+1}d_{m+1}}}{1 + \Gamma^{>}_m R^{>}_{m+1} e^{-j2k_{m+1}d_{m+1}}}, \quad \text{(A13)}$$

where $R^{>}_{m+1}$ and $T^{>}_i$ can be calculated by successive applications of (A5) and (A9), starting from region N. The superscript $(>)$ explicitly used to imply that the incident field travels in $(+\hat{z})$-direction.

All equations derived here are applicable for the case in which the incident field travels in $(-\hat{z})$-direction, provided $k_m \to -k_m$ and $(m \mp) \to (m \pm)$. Hence; the effective reflection and transmission coefficients for the geometry depicted in Fig. 13 are respectively defined as

$$R^{<}_m(z_m) = \frac{\Gamma^{<}_m + R^{<}_{m-1}e^{-j2k_{m-1}d_{m-1}}}{1 + \Gamma^{<}_m R^{<}_{m-1} e^{-j2k_{m-1}d_{m-1}}};$$

$$\Gamma^{<}_m = \frac{\eta_{m-1} - \eta_m}{\eta_{m-1} + \eta_m}, \quad \text{(A14)}$$

$$T^{<}_m(z_m) = \frac{\tau^{<}_m}{1 + \Gamma^{<}_m R^{<}_{m-1} e^{-j2k_{m-1}d_{m-1}}};$$

$$\tau^{<}_m = 1 + \Gamma^{<}_m. \quad \text{(A15)}$$

and

$$\Upsilon^{<}_{m,0} = \prod_{i=0}^{m+1} T^{<}_i(z_i)e^{-jk_{i-1}d_{i-1}}. \quad \text{(A16)}$$

Note that in this case the subscript m of the indices of the layers in Fig. 13 is monotonically decreasing; (i.e., $m \leq 0$; $z_{m-1} < z_m$).

It is evident from the above analysis that once the incident waves on either side of the source in region (0) are known, the voltages and currents of other regions of the transmission line will be specified.

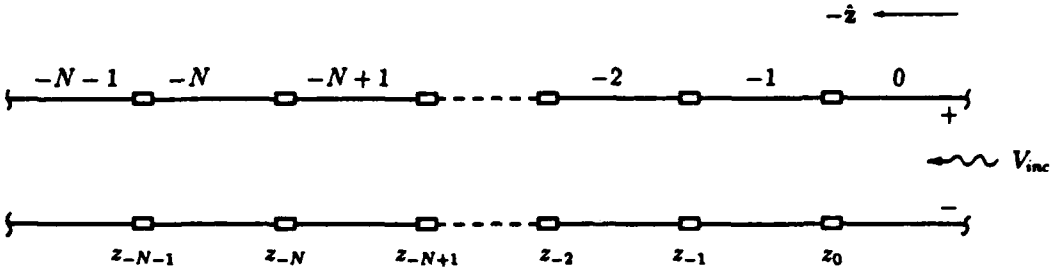


Fig. 13. General piecewise uniform transmission line; incident waves travel in $-\hat{z}$-direction.

REFERENCES

[1] A. Sommerfeld, *Partial Differential Equations in Physics*. New York: Academic, 1949.

[2] J. R. Wait, *Antenna Theory*, part 2, Collin and Zucker Eds., New York: McGraw-Hill, 1969, pp. 438-506.

[3] A. Banos, *Dipole Radiation in the Presence of a Conducting Half Space*. New York: Pergamon, 1966.

[4] C. H. Stoyer, "Electromagnetic fields of dipoles in stratified media," *IEEE Trans. Antennas Propag.*, pp. 547-552, July 1977.

[5] J. R. Wait, *Electromagnetic Wave Theory*. New York: Harper & Row, 1985.

[6] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*. Englewood Cliffs, NJ: Prentice-Hall, 1973, pp. 442-630.

[7] J. A. Kong, Ed., *Research Topics in Electromagnetic Wave Theory*. New York: Wiley-Interscience, 1981, pp. 210-234.

[8] J. A. Kong, *Theory of Electromagnetic Waves*, 2nd ed., New York: Wiley, 1986.

[9] S. Barkeshli and P. H. Pathak, "On the dyadic Green's function of the multilayered dielectric/ferrite media," in *Proc. IEEE Int. AP-S Symp. and National Radio Science Meeting*, Blacksburg, VA, June 15-19, 1987.

[10] ——, "On the dyadic Green's function of an infinite periodic array of arbitrary oriented dipole in a multi-layered media," in *Proc. IEEE Int. AP-S Symp. and National Radio Science Meeting*, Blacksburg, VA, June 15-19, 1987.

[11] J. S. Bagby and D. P. Nyquist, "Dyadic Green's function for integral electronic and optical circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, no. 2, pp. 206-210, Feb. 1987.

[12] M. S. Viola and D. P. Nyquist, "An observation on the Sommerfeld-integral representation of the electric dyadic Green's function for layered media," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-36, pp. 1289-1292, Aug. 1988.

[13] R. E. Collin, *Field Theory of Guided Waves*. New York: McGraw-Hill, 1960.

[14] P. H. Pathak, "On the eigenfunction expansion of electromagnetic dyadic Green's functions," *IEEE Trans. Antennas Propagat.*, vol. AP-31, pp. 837-846, 1983.

[15] C. T. Tai, *Dyadic Green's Function in Electromagnetic Theory*. Scranton, PA: Intext, 1971.

[16] J. Van Bladel, *Electromagnetic Fields*. New York: McGraw-Hill, 1964, pp. 537-545.

[17] J. R. Wait, *Geo-Electromagnetism*. New York: Academic, 1982.

[18] B. A. Munk, G. A. Burrell, T. W. Kornbau, "A general theory of periodic surfaces in stratified dielectric media," Tech. Rep. 784346-2, ElectroScience Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, Nov. 1977.

[19] B. A. Munk, "A general theory of periodic surfaces in a Stratified dielectric media," Tech. Rep. 715582-4, ElectroScience Laboratory, Dept. of Electrical Engineering, The Ohio State University, Columbus, Feb., 1986.

[20] S. Barkeshli and P. H. Pathak, "Radial propagation and steepest descent path representation of the planar microstrip dyadic Green's function," *Radio Science*, vol. 25, no. 2, pp. 161-174, Mar.-Apr. 1990.

[21] W. C. Chew, *Waves and Fields in Inhomogeneous Media*. New York: Van Nostrand Reinhold, 1990.

Sina Barkeshli (M'88) was born in Tehran, Iran, on July 30, 1956. He attended Shiraz University in Shiraz, Iran. He received the B.S.E.E. and M.S.E.E. from the University of Kansas in 1979 and 1982, respectively, and Ph.D. degree in electrical engineering from The Ohio State University in 1988.

From 1980 to 1982 he was with the Remote Sensing Laboratory of the University of Kansas as a Graduate Research Associate working on the problem of sea surface scattering and radar signal processing for the synthetic aperture radar (SAR). From 1982 until 1988 he worked as a Graduate Research Associate at The Ohio State University Electroscience Laboratory conducting research on radar target identification, geometrical theory of diffraction, and microstrip antennas. Currently he is a Senior Member of Technical Staff at Sabbagh Associates, Inc. in Bloomington, Indiana. His current interests are primarily focused on analytical and computational electromagnetics, nonlinear inversion algorithms, interactions of electromagnetic waves with complex media and charged particles, electromagnetic wave modeling of microstrip, millimeter wave and optical systems, as well as signal and image processing.

Dr. Barkeshli is a member of Sigma Xi and Eta Kappa Nu.

Prabhakar H. Pathak (M'76-SM'81-F'86) received the B.Sc. degree in physics from the University of Bombay, India, in 1962, the B.S. degree in electrical engineering, from Louisiana State University, Baton Rouge, in 1965, and the M.S. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, in 1970 and 1973, respectively.

From 1965 to 1966 he was an instructor in the Department of Electrical Engineering at the University of Mississippi, Oxford. During the summer of 1966, he worked as an electronics engineer with the Boeing Company in Renton, WA. Since 1968 he has been with the Ohio State University ElectroScience Laboratory, where his research interests have centered on mathematical methods, electromagnetic antenna and scattering problems, and uniform ray techniques. He is also an Associate Professor in the Department of Electrical Engineering at Ohio State University, where he teaches courses in electromagnetics, antennas, and linear systems.

Dr. Pathak has participated in invited lectures and in several short courses on the uniform geometrical theory of diffraction, both in the United States and abroad. He has also authored and coauthored chapters on ray methods for five books. Dr. Pathak is a member of Commission B of the International Scientific Ratio Union (URSI), and of Sigma Xi.

Improving the Performance of a Slotted ALOHA Packet Radio Network with an Adaptive Array

James Ward, *Member, IEEE,* and R. T. Compton, Jr., *Fellow, IEEE*

Abstract—**The use of an adaptive antenna array is presented as a means to improve the performance of a slotted ALOHA packet radio network. An adaptive array creates a strong capture effect at a packet radio terminal by automatically steering the receive antenna pattern toward one packet and nulling other contending packets in a slot. A special code preamble and randomized arrival times within each slot allow the adaptive array to lock onto one packet in each slot. The throughput and delay performance of a network with an adaptive array is computed by applying the standard Markov chain analysis of slotted ALOHA [1], [2]. It is shown that throughput levels comparable to CSMA are attainable with an adaptive array without the need for stations to be able to hear each other. The performance depends primarily on the number of adaptive array nulls, the array resolution, and the length of the randomization interval within each slot.**

I. Introduction

ALOHA packet radio communication systems are of interest because they provide a simple way of multiplexing many users into a single radio channel. In these systems radio terminals transmit packets to each other whenever they have information to send, regardless of whether other terminals may be transmitting at the same time. Because terminals do not coordinate their transmissions, packets from different terminals frequently collide. A collision destroys all packets involved, and these packets must then be retransmitted after a random delay. Collisions limit the maximum throughput at one receiver in an ALOHA system to 18% if the system is unslotted and to 36% if it is slotted [3].

Because of these low throughputs, much effort has been devoted to finding improved packet radio protocols. One well-known improvement is carrier sense multiple access (CSMA) [4], in which terminals listen to the channel before transmitting to determine if it is busy. If the channel is busy, transmission is delayed until the channel becomes idle. Kleinrock and Tobagi have shown that choosing the retransmission probability carefully in a CSMA system can yield high throughputs [4]. However, the usefulness of CSMA depends on whether all terminals in the network can hear one another. When this is not the case, as in satellite or mobile communications, CSMA is less effective.

In the standard slotted ALOHA analysis, it is assumed that if two or more packets arrive in the same slot, none of them is received correctly. In reality, the correct reception of a packet depends not only on whether interfering packets are present, but also on the received power of each packet. Roberts [5] first noted that if one of the packets is of much higher received power than the others, it may still be correctly received. This "power capture" effect improves the throughput and delay performance of a packet radio system. Power capture has been studied by Abramson [3] and Namislo [7] when it occurs naturally as a result of different propagation distances from transmitter to receiver and/or channel fading. Lee [6] considered assigning random signal levels to the stations to induce the capture effect. Also, since the received power from a given direction is proportional to the receiver antenna response in that direction, directional antennas can be used to create the capture effect at the receiver. Binder *et al.* [8] have considered using directional antennas to resolve potential crosslink conflicts in a multiple satellite packet system. In their work the direction to which an antenna is steered is obtained *a priori* from a form of scheduling used to set up each communication link. Their scheduling procedure, in addition to providing direction information, also reduces the contention somewhat at the expense of increased packet delay.

In this paper, we examine the use of an *adaptive* antenna array to create a capture effect and thus improve the performance of a slotted ALOHA system. An adaptive array is an antenna system that controls its own pattern in response to the signal environment [9], [10]. An adaptive array can capture a packet by pointing the peak antenna response toward that packet while simultaneously forming pattern nulls on other interfering packets [11]. An adaptive array can do this automatically without requiring any *a priori* direction information. Thus, there is no need for prearranged scheduling in a system with an adaptive array and the delay performance should be improved. Furthermore, an adaptive array provides a much stronger capture effect than an ordinary directional antenna, because pattern nulls are placed in the directions of contending packets. We shall show that the use of an adaptive array can provide throughput and delay performance comparable to that of CSMA. Moreover, with an adaptive array there is no need for users to be able to hear each other.

In Section II we describe the communication system we shall consider. Section III gives a brief overview of adaptive arrays. Section IV describes how an adaptive array can acquire

Paper approved by the Editor for CATV of the IEEE Communications Society. Manuscript received February 10, 1990. This work was supported in part by the U.S. Army Research Office, Research Triangle Park, NC, and by the Office of Naval Research, Arlington, VA, under Contracts DAAL03-89-K-0073 and N00014-89-J-1007 with The Ohio State University Research Foundation, Columbus, OH.

J. Ward was with the ElectroScience Laboratory, The Ohio State University. He is now with M.I.T. Lincoln Laboratory, Lexington, MA, 02173.

R. T. Compton, Jr., is with the ElectroScience Laboratory, The Ohio State University, Columbus, OH 43212.

IEEE Log Number 9106306.

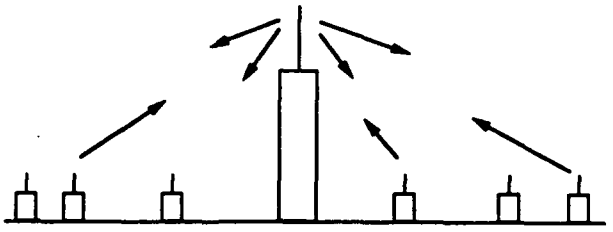


Fig. 1. A single-hop packet radio system.

the first packet to arrive in a slot while nulling subsequent packets in that slot. In Section V we calculate the throughput and delay performance of a packet system using an adaptive array. Section VI presents numerical results, and Section VII contains our conclusions.

II. The Communication System Model

We consider a simple ALOHA system in which a repeater links a network of radio terminals, as shown in Fig. 1. In this network terminals transmit messages to each other through the repeater. We assume time is slotted and that the network uses a slotted ALOHA packet radio protocol. Transmissions between terminals occur randomly in each time slot. Each terminal transmits a packet in a given slot whenever it has one to send, without regard for whether other terminals may be transmitting in that same slot.

All packets are transmitted to the central repeater, which retransmits them back to the network. The repeater is assumed to be a *store-and-forward* repeater. It demodulates each packet and checks it for errors. If there are no errors, the packet is retransmitted on the downlink. If there are errors, the packet is discarded. The repeater downlink is on a different frequency than the uplink, so both the repeater and the local terminals can transmit and receive at the same time. Since only the repeater transmits on the downlink, there is no contention on the downlink.

Each terminal monitors all downlink packets. By examining the address contained in each packet, a terminal determines whether it is the intended recipient of that packet. A terminal retains packets addressed to itself and discards others. Moreover, when a terminal transmits a packet of its own over the repeater, it listens for that packet on the downlink to determine if the packet was successfully forwarded. If the packet is not heard on the downlink, it is assumed that the packet suffered a collision on the uplink, and the packet is retransmitted after a delay of some random number of slots.

We assume the receiving antenna at the repeater is an adaptive array.[1] The purpose of the adaptive array is to aim the repeater antenna pattern at the first packet to arrive in each slot and then to null subsequent interfering packets in that slot, to prevent them from destroying the first packet. This technique will allow one packet to be received successfully, even when several packets arrive in the same slot. (In a conventional ALOHA system, all packets are destroyed when a collision occurs.) The method used to form the antenna pattern is described in Section IV below.

[1] The transmitting antenna at the repeater is assumed to cover all the users of the network so that each terminal can hear all downlink packets.

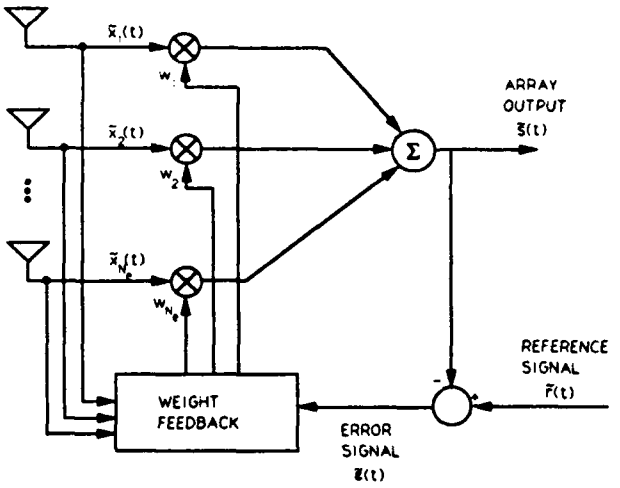


Fig. 2. An adaptive array.

Now let us consider this system in more detail. We begin in the next section by reviewing the adaptive array concepts needed.

III. Adaptive Arrays

An adaptive array is an antenna system that controls its own pattern, by means of feedback, while the antenna operates [9], [12], [13]. The signal from each element in an adaptive array is multiplied by a weight and then summed to produce the array output signal. A control system adjusts the weights to maximize the signal-to-interference-plus-noise-ratio (SINR) at the array output. After adapting, the pattern of an adaptive array has a beam pointed at the desired signal and has nulls on interfering signals. In a packet radio system, the desired signal is just the first packet in each slot. The interfering signals are the other packets contending for channel access in that slot.

Fig. 2 shows an adaptive array with N_e elements. The signal $\tilde{x}_j(t)$ from element j is multiplied by a weight w_j and then summed to produce the array output signal $\tilde{s}(t)$. The weights are controlled by a feedback system that minimizes the mean-square value of the *error signal* $\tilde{\epsilon}(t)$, which is the difference between the array output $\tilde{s}(t)$ and a signal $\tilde{r}(t)$ called the *reference signal.* The reference signal is a locally generated signal that determines which received signals are retained in the array output and which are nulled. Minimizing the mean-square value of $\tilde{\epsilon}(t)$ is equivalent (for narrow-band signals) to maximizing the signal-to-interference-plus-noise ratio (SINR) at the array output and causes the array to steer a beam toward any signal correlated with the reference signal and to null any signal uncorrelated with it [9].

It may be shown [9] that the optimal (maximum SINR) array weights are given by

$$\boldsymbol{W} = \boldsymbol{\Phi}^{-1}\boldsymbol{S} \tag{1}$$

where $\boldsymbol{W}$ is the weight vector,

$$\boldsymbol{W} = [w_1, w_2, \cdots, w_{N_e}]^T, \tag{2}$$

Φ is the covariance matrix,

$$\Phi = E[X^* X^T], \tag{3}$$

and S is the *reference correlation vector,*

$$S = E[X^* \tilde{r}(t)]. \tag{4}$$

In these equations, X is the signal vector, i.e., a vector containing the element signals,

$$X = [\tilde{x}_1(t), \tilde{x}_2(t), \cdots, \tilde{x}_{N_e}(t)]^T, \tag{5}$$

$E[\,]$ denotes expectation, * denotes complex conjugate, and T denotes transpose. The weights in (1) are known as the Wiener weights.

A well-known method of controlling the weights in an adaptive array is the sample matrix inverse technique of Reed, Mallett, and Brennan [14]. In this technique, the element signals are sampled periodically in I and Q (inphase and quadrature) channels and an estimate of the covariance matrix is computed from the sampled signals. If $X(j)$ denotes the value of the signal vector X at sample time j, the *sample covariance matrix* is computed from

$$\hat{\Phi} = \sum_{j=1}^{K} X^*(j) X^T(j) \tag{6}$$

where K is the number of samples used. The notation $\hat{\Phi}$ is used to indicate that (6) is an estimate of Φ in (1). The *sample reference correlation vector* $\hat{S}$ is computed from

$$\hat{S} = \sum_{j=1}^{K} X^*(j) \tilde{r}(j) \tag{7}$$

where $\tilde{r}(j)$ is sample j of the reference signal $\tilde{r}(t)$. The optimal weights are then estimated by solving the system of equations

$$\hat{\Phi} W = \hat{S} \tag{8}$$

for the weight vector. Reed *et al.* [14] have shown that this technique produces an average SINR within 3 dB of the optimal SINR if the samples $X(j)$ are statistically independent and if the number of samples K is approximately twice the number of array elements.

When several signals are incident on the array, the reference signal $\tilde{r}(t)$ determines which signals are retained in the array output and which are nulled. Any signal correlated with $\tilde{r}(t)$ is retained in the array output and any signal uncorrelated with $\tilde{r}(t)$ is nulled [9]. To use an adaptive array in a communication system, the main challenge is to find a way to obtain a reference signal correlated with the desired signal and uncorrelated with the interference. In Section IV we describe a method for doing this with packets.

An adaptive array has two limitations that are important for this application. The first is that an array with N_e elements has only $N_e - 1$ degrees of freedom in its pattern [9]. Each null or beam maximum formed by the array requires one degree of freedom. In our case, the array needs to form a beam maximum on one packet and nulls on all other packets in a slot. Thus, an N_e-element array using one degree of freedom to form the required beam maximum can also form nulls on up to $N = N_e - 2$ packets. When there are more interfering signals than the available degrees of freedom, the array will not be able to null them all [9].

Another limitation of adaptive arrays is that a given array has only a certain ability to resolve signals in space. If the arrival angles of an interfering packet and the desired packet are too close, the array cannot simultaneously null the interference and form a beam on the desired packet. In this case, the array output desired signal-to-noise ratio drops and the adaptive array may not capture the desired packet. To characterize the resolution capability of an adaptive antenna, we define the *resolution width* θ_r to be the minimum angular separation between two signals at which the adaptive array can place a pattern maximum on one signal and null the other. The resolution width θ_r is taken to be $\theta_b/2$, where θ_b is the beamwidth of the array, i.e., the angular separation between the first nulls on each side of the mainbeam. θ_b depends primarily on the array aperture size but also to a lesser extent on the element patterns and the number of elements. In the analysis below, we relate the performance of the packet radio system to the number of nulls available and to the resolution capability of the array.

With this background, we now describe a technique for operating an adaptive array in a packet radio system.

IV. Acquisition

The main difficulty in using an adaptive array in a packet radio system is the acquisition problem, i.e., the problem of forming the beam on the first packet and nulling subsequent packets in the slot. Each packet to be received by the array will arrive at an unknown time and from an unknown direction. The array must form its pattern on a packet very rapidly, in time to receive the message portion of the packet. To allow an adaptive array to do this, we add a special two-part preamble to the beginning of the packet. The first part of this preamble will be used to trigger the acquisition process, and the second part will be used to form the array pattern on the packet.

Fig. 3 shows the organization of a packet. A packet will be formed by first adding an address preamble to the beginning of a fixed number of message bits, as shown in the top of Fig. 3. The address preamble will identify the destination terminal and may contain other information such as the originating terminal or a packet number. Next, the combined address and message segments will be encoded with an (n, k) linear block code [15], which will be used for error detection at the repeater. Finally, after encoding, an additional two-part preamble will be added to the beginning of the packet. This preamble, called the *acquisition preamble,* will be used to lock the array pattern on the packet.

The acquisition preamble will consist of two consecutive code sequences, called Codes 1 and 2. Code 1 will be a 13 bit Barker code [16], which has a highly peaked aperiodic autocorrelation function as shown in Fig. 4. Code 2 will be one or more periods of a pseudonoise (PN) code [17]. The periodic autocorrelation function of such a code has a sharp

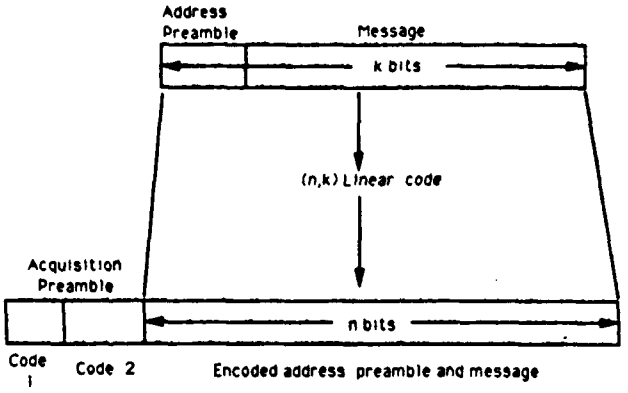


Fig. 3. Packet organization.

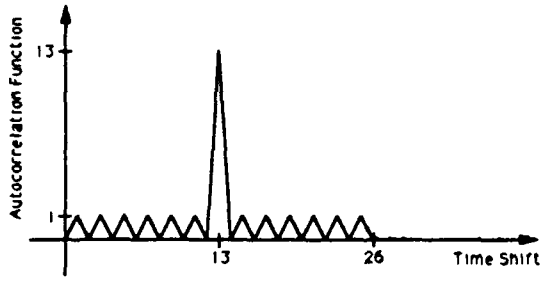


Fig. 4. Autocorrelation function of a 13 bit Barker code.

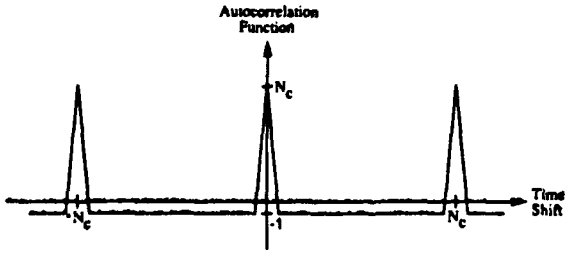


Fig. 5. Autocorrelation function of a PN code.

peak of height N_c at zero shift (and at shifts of any multiple of the code period) and then drops to a constant value of -1 for shifts over 1 bit where N_c is the code period, as shown in Fig. 5.

To allow the packet acquisition, the width of the slot T_s will be made larger than the packet width T_p by an *uncertainty interval* T_u, as shown in Fig. 6. To exploit the autocorrelation properties of the preamble codes, the starting times of packet transmissions from all terminals will be randomized over the interval T_u, as in [18]. The uncertainty interval also makes the acquisition process fair (by preventing stations closest to the repeater from always acquiring the repeater first) and gives the designer control over the probability that two packets arrive at almost the same instant.

The adaptive array will operate as follows. At the beginning of each slot, when the repeater is ready to acquire a new packet, the array weights will be set so the array pattern covers all users in the net. Such a pattern is easily obtained by turning one array weight on and the rest off. With one weight on, the

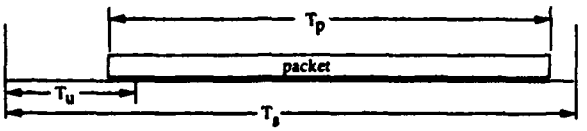


Fig. 6. Slot width, packet width, and uncertainty interval.

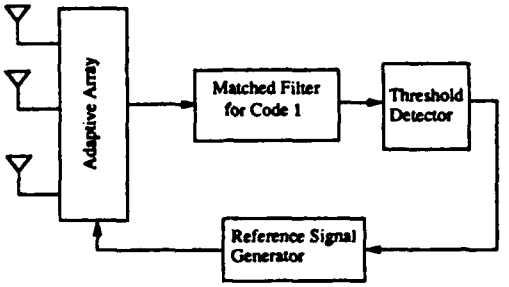


Fig. 7. Packet acquisition circuitry.

array pattern is just the pattern of the element that is turned on. This element pattern will be chosen so it covers the entire net. We call this the *uniform coverage* mode. In this mode, any user can access the system.

To acquire an incoming packet, we use the following technique. At the array output is a filter matched to Code 1, followed by a threshold detector, and then a reference signal generation circuit, as shown in Fig. 7. Assume first that only one packet arrives during the slot. With the array in its uniform coverage mode, the incoming packet will pass through the array and into the matched filter. The output of this filter will contain a sharp peak at the end of Code 1. This peak will serve as a timing spike to trigger generation of a reference signal during Code 2. The reference signal will be a signal modulated by the same PN code as in Code 2. The timing spike will start the reference signal at the proper time so it is correlated with the received packet during Code 2. The reference signal will continue only during Code 2. The array pattern will be adapted during Code 2. Because the reference signal code is synchronized with Code 2 in the packet, the array will optimize its weights for reception of the packet.[2] At the end of Code 2, the array weights will be frozen. The array pattern will then be held fixed during the address and message portions of the packet.

Now suppose two or more packets are received in the same slot. Each of these packets will cause a timing spike at the matched filter output. But only the first timing spike will trigger reference signal generation and begin array adaptation. Timing spikes due to later packets will be ignored by the system, because the acquisition circuit will be designed so that once it has been triggered, it will not trigger again in the same slot.

Because the reference signal code will be aligned with Code 2 of the first packet, it will be essentially uncorrelated

[2]The reference signal does not have to be locked in frequency or phase to the received packet for this process to work. The only requirements are that the PN codes be synchronized to within about one fourth of a code bit, and that the difference between the reference signal frequency and the received signal frequency be less than the reciprocal of the adaptation time [19], [20].

with the second packet *as long as the second packet is at least one bit later than the first.* This is so because the autocorrelation function of a PN code has a very low value for shifts of 1 bit or more. (See Fig. 5.) The second packet and all later packets will therefore be regarded as interference by the adaptive array and will be nulled. At the end of Code 2, the array pattern will be optimized for receiving the first packet and will have nulls on later packets.

If the second packet arrives *less than one bit after the first,* the first two packets will be correlated. The adaptive array will not null the second packet in this case and there will be no throughput. In this case we say that the first packet is not acquired. With the uncertainty interval T_u properly chosen, however, the probability of this event is small. The throughput analysis below takes this possibility into account.

The uncertainty interval T_u and the durations of Codes 1 and 2 will be chosen so that all packets in a given slot begin no later than during the Code 2 preamble of the first packet in the slot. For this reason it is possible to finish adapting the array weights at the end of Code 2 and fix the array pattern during the address and message segments. The adapted pattern at the end of Code 2 will have nulls on the interfering packets, and these will be retained for the rest of the slot.

In the analysis below, we assume that the packet SNR is high enough so that if a packet is present, it is always detected by the acquisition circuitry. We also assume that the possibility of a false alarm, i.e., the triggering of a reference signal without the presence of a corresponding packet, is negligible. We assume the array acquires the first packet to arrive in each slot as long as another packet does not arrive in that slot less than one bit after the first. However, even if a packet is acquired, it may still not be successful. An acquired packet will be unsuccessful in either of two cases:

1) when more interfering packets arrive during a slot than the number of available nulls, or
2) when another packet arrives too close in angle to the acquired packet.

At the end of each slot, the array is reset into its uniform coverage mode, and the acquisition cycle starts over for the next slot.

We now consider the throughput and delay performance of a packet radio repeater using an adaptive array with this acquisition technique.

V. Throughput and Delay Analysis

To determine the throughput and delay performance, we apply the Markov chain analysis of a slotted ALOHA network [1], [2] to include the effects of the adaptive array and the acquisition process.

We consider a finite population of M terminals transmitting to a central repeater equipped with an adaptive array. At the beginning of each slot, each terminal is either *blocked* or *unblocked,* depending on whether its previously transmitted packet was unsuccessful or successful. An unblocked terminal transmits a packet with probability p_n in a slot. Only unblocked terminals generate new packets. A blocked terminal retransmits its backlogged packet with probability p_r in each slot until successful, at which time it becomes unblocked and resumes transmitting new packets. Typically, $p_r > p_n$ so that backlogged packets are quickly cleared. At the end of each slot, the downlink transmission provides immediate feedback to the terminals regarding the success of their packets.

Let X_k denote the number of blocked terminals at the beginning of slot k. The number of blocked terminals at the end of the slot depends only on the number at the beginning of the slot and the events occurring during the slot. Thus, the time-varying state of the network can be described by a Markov chain, where the state represents the number of blocked terminals. At slot k, the state X_k can vary between 0 and M. We shall compute the one-step transition matrix $\boldsymbol{P} = [P_{i,j}]$ and then the equilibrium probabilities of the Markov chain describing this system.

In a given slot, there will be a total of $n_t = n_n + n_r$ packets transmitted where n_n and n_r are the number of new and previously backlogged packets transmitted in the slot. Given the state $X_k = i$, n_n and n_r are independent Bernoulli random variables with distributions

$$Q_n(l|i) \triangleq \Pr\{n_n = l | X_k = i\} = \binom{M-i}{l} p_n^l (1-p_n)^{M-i-l} \quad (9)$$

$$Q_r(l|i) \triangleq \Pr\{n_r = l | X_k = i\} = \binom{i}{l} p_r^l (1-p_r)^{i-l}. \quad (10)$$

Thus, the distribution of the total number of packets per slot is

$$Q_t(l|i) \triangleq \Pr\{n_t = l | X_k = i\} = \sum_{s=0}^{l} Q_n(s|i) Q_r(l-s|i). \quad (11)$$

Let $P_s(l)$ be the probability that a packet is *successful* given that l packets are transmitted in the slot. The success probabilities $P_s(l)$, which depend on the adaptive array characteristics and the acquisition parameters, will be determined below. Given $P_s(l)$, the transition probabilities $P_{i,j}$ may be found by enumerating the possible ways that each transition may occur.

- $j < i-1, i = 2, \cdots, M$: Not possible, since at most one backlogged packet can be cleared in a slot.

$$P_{ij} = 0. \quad (12)$$

- $j = i-1, i = 1, \cdots, M$:
  1) $n_n = 0, n_r \geq 1$, and one backlogged packet is successful.

$$P_{i,i-1} = Q_n(0|i) \sum_{l=1}^{i} Q_r(l|i) P_s(l). \quad (13)$$

- $j = i+k, i = 0, \cdots, M,\ k = 0, \cdots, M-i$:
  1) $n_n = k+1, n_r \geq 0$, and one packet is successful.

2) $n_n = k, n_r \geq 0$, and none of the transmitted packets are successful.

$$P_{i,i+k} = Q_n(k+1|i)\sum_{l=0}^{i} Q_r(l|i)P_s(l+k+1) + Q_n(k|i)\sum_{l=0}^{i} Q_r(l|i)(1-P_s(l+k)). \tag{14}$$

This Markov chain analysis is similar to that of Namislo [7]. (Namislo determines the success probabilities for a fading environment by using a Monte–Carlo simulation. We will derive them directly for the adaptive array.)

To compute the $P_s(l)$, we first note the distinction between *acquired* packets and *successful* packets. An acquired packet is one for which the array acquisition circuitry generates a reference signal that is not correlated with any other packets. Note that for a packet to be successful, it must first be acquired by the array. Once a packet is acquired, it is successful only if the adaptive array can form a beam on the acquired packet and place pattern nulls in the directions of the other contending packets.

Given that there are l packets in a slot, we characterize each packet by an arrival *time* t_i, $i = 1, \cdots, l$ within a slot and an arrival *angle* $\theta_i, i = 1, \cdots, l$. In accordance with the acquisition procedure in Section IV, we assume that the t_i are i.i.d. random variables uniformly distributed on the uncertainty interval $[0, T_u]$ within the slot. We also assume packet arrival angles are i.i.d. random variables (independent of the arrival times) uniformly distributed in azimuth $[0, 2\pi]$ about the central repeater node. Then

$$P_s(l) = P_a(l)P_{s|a}(l) \tag{15}$$

where $P_a(l)$ is the probability that a packet is acquired given l packets are incident, and $P_{s|a}(l)$ is the probability that a packet is successful given it is acquired and l packets are present in the slot. The $P_a(l)$ depend on the arrival times and the length of the uncertainty interval, while the $P_{s|a}(l)$ depend on the arrival angles, the resolution capability of the adaptive array, and the number of available nulls.

With the preamble code structure described in Section IV, the first packet in a slot is acquired as long as all subsequent packets in that slot arrive at least one bit duration T_b later than the first packet. If the first packet is not acquired, no packets are acquired for that slot. Thus,

$$P_a(l) = l\Pr\{t_2 > t_1 + T_b, t_3 > t_1 + T_b, \cdots, t_l > t_1 + T_b\} \tag{16}$$

where the factor of l accounts for the fact that any of the l packets transmitted can be the first packet in the slot. If only a single packet is transmitted in a slot, it is acquired, so

$$P_a(1) = 1. \tag{17}$$

For, $l \geq 2$, we use the uniform distribution of the transmission times to write

$$P_a(l) = l\int_0^{T_u - T_b}\int_{t_1+T_b}^{T_u}\cdots\int_{T_1+T_b}^{T_u}\left(\frac{1}{T_u}\right)^l dt_l \cdots dt_2\, dt_1. \tag{18}$$

Thus, from (17) and (18),

$$P_a(l) = \begin{cases} 1; & l = 1 \\ \left(1 - \frac{1}{u}\right)^l; & l > 1 \end{cases} \tag{19}$$

where $u = T_u/T_b$ is the length of the uncertainty interval in bits.

Once a packet is acquired, two conditions must be satisfied for it to be successful. First, there must be no more than $N = N_e - 2$ additional packets transmitted in the slot, because the adaptive array can place pattern nulls in at most N directions. Second, no other packet can arrive from an angle within $\theta_b/2$ of the acquired packet arrival angle. If this happens, the adaptive array will be unable to resolve the acquired and interfering packets and there will be no throughput for the slot.

The $P_{s|a}(l)$ may be computed as follows. First, we have

$$P_{s|a}(1) = 1 \tag{20}$$

since with only one packet present there are no other packets to interfere with the acquired packet. Moreover, because the adaptive array has only N nulls, we set

$$P_{s|a}(l) = 0, \qquad l > N + 1. \tag{21}$$

To find $P_{s|a}(l)$ for $2 \leq l \leq N+1$, recall that θ_1 is the arrival angle of the acquired packet and define $D_1 = [\theta_1 - \theta_b/2, \theta_1 + \theta_b/2]$. Then

$$\begin{aligned} P_{s|a}(l) &= \Pr\{\theta_2 \notin D_1, \theta_3 \notin D_1, \cdots, \theta_l \notin D_1\} \\ &= E_{\theta_1}\{\Pr\{\theta_2 \notin D_1, \theta_3 \notin D_1, \cdots, \theta_l \notin D_1|\theta_1\}\} \\ &= E_{\theta_1}\left[\prod_{i=2}^{l}\Pr\{\theta_i \notin D_1|\theta_1\}\right] \end{aligned} \tag{22}$$

where $E_{\theta_1}[]$ denotes an expectation over the random variable θ_1, and we have taken advantage of the independence of the arrival angles.

However,

$$\Pr\{\theta_i \notin D_1|\theta_1\} = \left(1 - \frac{\theta_b}{2\pi}\right), \tag{23}$$

which is independent of θ_1. Thus, (22) becomes

$$P_{s|a}(l) = \left(1 - \frac{\theta_b}{2\pi}\right)^{l-1}, \quad 2 \leq j \leq N+1. \tag{24}$$

Hence, from (15), (19), and (24), the success probabilities are

$$P_s(l) = \begin{cases} 0; & l = 0 \\ 1; & l = 1 \\ \left(1 - \frac{1}{u}\right)^l\left(1 - \frac{\theta_b}{2\pi}\right)^{l-1}; & 2 \leq l \leq N+1 \\ 0; & l > N+1 \end{cases} \tag{25}$$

Given that the system is in state j, the probability of a successful packet transmission is the conditional throughput $S(j)$, given by

$$S(j) = \sum_{l=1}^{M} Q_t(l|j) P_s(l). \tag{26}$$

The average number of *new* packets entering the system in state j is

$$S_{in}(j) = (M - j)p_n. \tag{27}$$

The Markov chain described above is irreducible. Since we assumed a finite population, all states are recurrent non-null. The states are also aperiodic. Consequently, this Markov chain has a limiting distribution denoted by

$$\boldsymbol{\pi} = [\pi(0), \pi(1), \cdots, \pi(M)] \tag{28}$$

where

$$\pi(j) = \Pr\{X_\infty = j\} = \lim_{n\to\infty} \Pr\{X_{k+n} = j | X_k = i\}. \tag{29}$$

The steady-state probabilities are found by solving the linear system of equations [21]

$$\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{P} \tag{30}$$

along with the constraint that

$$\sum_{j=0}^{M} \pi(j) = 1. \tag{31}$$

Once the $\pi(j)$ are found, they can be used to determine the average throughput, delay, and backlog of the system.

Given $\pi(j)$, the average number of blocked terminals $\overline{B}$ is

$$\overline{B} = \sum_{j=0}^{M} j\pi(j). \tag{32}$$

and the average throughput is

$$\overline{S} = \sum_{j=0}^{M} S(j)\pi(j). \tag{33}$$

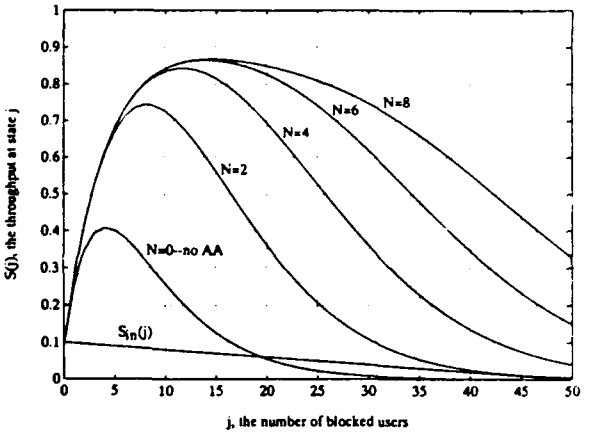
In the steady state, the average input rate equals the average throughput, so

$$\overline{S}_{\text{in}} = S_{\text{in}}(\overline{B}) = \overline{S}. \tag{34}$$

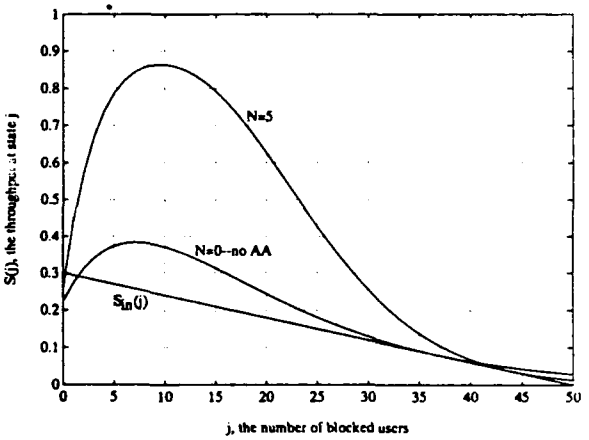
We use Little's theorem [23] to express the average delay $\overline{D}$ experienced by a new packet as

$$\overline{D} = \frac{\overline{B}}{\overline{S}_{\text{in}}} = \frac{\overline{B}}{\overline{S}}. \tag{35}$$

We now use these results to examine the performance of a slotted ALOHA system with an adaptive array.



(a)



(b)

Fig. 8. Conditional throughput comparison. For the curves with an adaptive array: $\theta_b = 10°$, $u = 62$. (a) $M = 50$, $p_n = 0.002$, $p_r = 0.2$. (b) $M = 50$, $p_n = 0.006$. Without the adaptive array, $p_r = 0.1$; $p_r = 0.2545$ with the adaptive array.

VI. Results

First we examine the conditional throughput $S(j)$ of systems with and without an adaptive array. We consider a network of 50 users. We start with an example where $p_n = 0.002$ and $p_r = 0.2$. For this case, $Mp_n = 0.1$, which is a low traffic situation where slotted ALOHA may typically be used. Fig. 8(a) shows the conditional throughput $S(j)$ and the new packet input rate $S_{\text{in}}(j)$ versus the state j. Curves for various numbers of adaptive array nulls are also shown. For these curves we have $\theta_b = 10°$ and $u = 62$. There is a significant increase in conditional throughput as the adaptive array is added and the number of nulls is increased. Also, note that there is a fixed number of nulls above which little further improvement is gained.

The stability problems of ALOHA systems have been well documented [1], [2], [22]. The finite population ALOHA model is said to be stable if there is a single intersection point of the $S(j)$ and $S_{\text{in}}(j)$ curves and this intersection point is in a region of low delay. In Fig. 8 we have intentionally chosen p_r high enough so that the system without an adaptive array is unstable. The curves with an adaptive array are stable. Moreover, for an adaptive array with 4 nulls or more, the

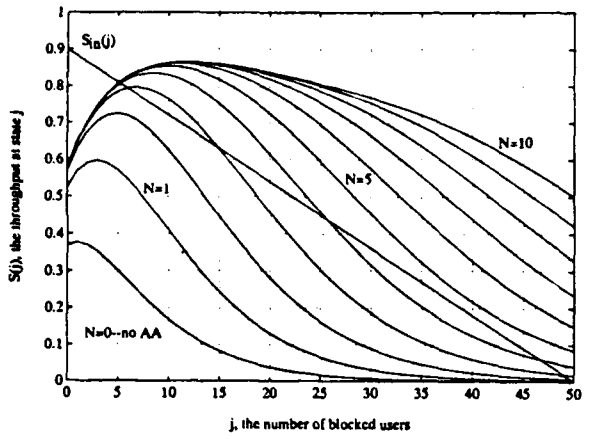


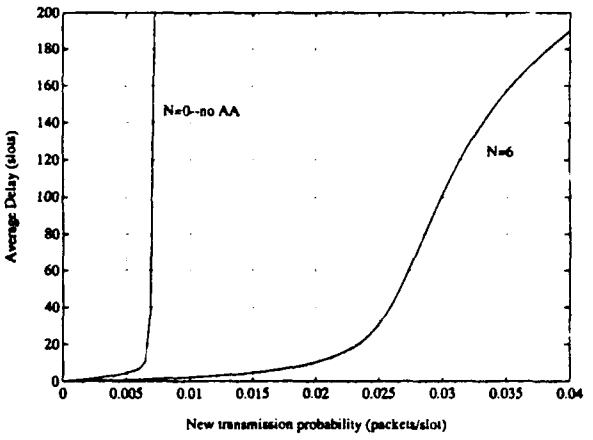
Fig. 9. Conditional throughput for a network of 50 users with $p_n = 0.018$, as the number of adaptive array nulls is varied. $\theta_b = 10°$ and $u = 62$.

value of p_n could be raised substantially without introducing instability. Thus, it is seen that the adaptive array has a stabilizing effect on the system.

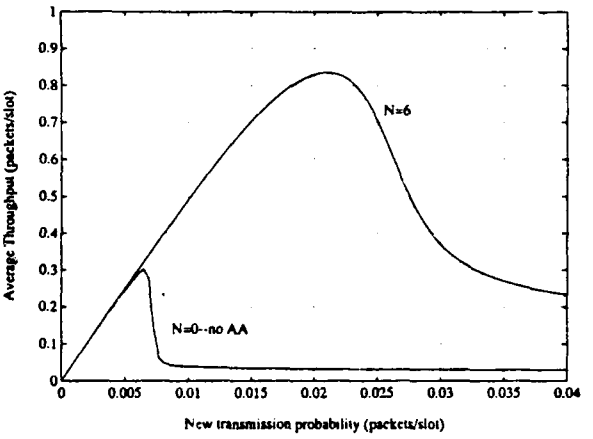
In Fig. 8(b) we compare two stable cases for $p_n = 0.006$. In each case we have chosen the largest retransmission probability possible for stable operation. Without the adaptive array, p_r is set to 0.1, which results in an average throughput of $\overline{S} = 0.288$ packets/slot, an average backlog of $\overline{B} = 2.01$ users, and an average delay of $\overline{D} = 6.99$ slots/packet. The maximum possible average throughput is $Mp_n = 0.3$ packets/slot. With an adaptive array, p_r is set to 0.2545, resulting in $\overline{S} = 0.299$, $\overline{B} = 0.241$, and $\overline{D} = 0.806$ slots. For such low traffic scenarios, the adaptive array provides only a slight increase in throughput but a marked improvement in the delay performance.

The main advantage of using an adaptive array in an ALOHA network is the ability to handle much higher traffic rates and operate at a much higher throughput than is possible in a standard ALOHA system. In Fig. 9, we consider a case with $p_n = 0.018$, so that on average, more than $Mp_n = 0.9$ packets (new plus backlogged) are transmitted per slot. We fix $p_r = 0.2$. To have a stable system, the adaptive array needs at least 5 nulls. For $N \geq 5$, the average throughput is 0.8 packets/slot. This example shows how performance can be improved by increasing the adaptive array capabilities. We note that a throughput of 0.8 is comparable to typical values attainable by CSMA [24], and with slotted ALOHA under other capture mechanisms [5]–[7].

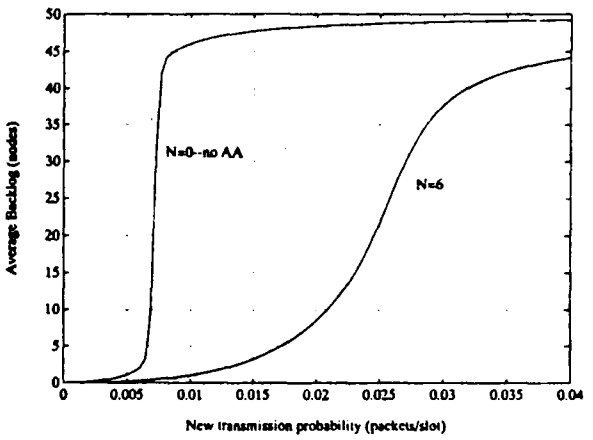
In general, performance improves as the number of adaptive array nulls increases or as the array beamwidth is reduced. Increasing the number of available nulls allows more collisions to occur without reducing the number of successful packets. Reducing the array beamwidth allows the array to successfully null interfering packets over a larger angular region. Performance is also improved as the length of the uncertainty interval is increased. (Of course, a longer uncertainty interval requires a longer slot width and reduces the number of message bits transmitted per unit time.) As the adaptive array capabilities (resolution, number of nulls) are increased, average throughputs close to unity can be approached. The limiting case of $\theta_b = 0°$, $u = \infty (T_b = 0)$, $N = M - 1$ corresponds to perfect capture where one packet is successful in every slot in which at least one packet is transmitted.



(a)



(b)



(c)

Fig. 10. Average $\overline{S}$, $\overline{B}$, $\overline{D}$ performance. Without the adaptive array, $p_r = 0.1$. With the adaptive array, $p_r = 0.2$, $\theta_b = 10°$, $u = 62$. (a) Average delay versus p_n. (b) Average throughput versus p_n. (c) Average backlog versus p_n.

Fig. 10 compares the average delay, throughput, and backlog performance of systems with and without an adaptive array for the case $\theta_b = 10°$, $u = 62$, and $N = 6$. The retransmission probability p_r is 0.1 without the adaptive array and 0.2 with the

array. These curves were obtained by varying p_n and computing $\overline{S}$, $\overline{D}$, and $\overline{B}$ from (32)–(35). We see from Fig. 10(a) that the delay with the adaptive array is always better than without it, and the difference is greater as p_n is increased (as the input traffic is increased). Fig. 10(b) shows the average throughput. For low traffic, both systems are stable and provide nearly the maximum possible throughput. However, the system without the adaptive array becomes unstable at relatively small p_n while the throughput with the adaptive array keeps increasing, to a maximum of near 0.83. Finally, if p_n is increased too far, the system with the adaptive array also becomes saturated and the network becomes highly backlogged. The average backlog for the two cases is shown in Fig. 10(c). Again, these curves indicate that by using an adaptive array, we can achieve acceptable delay at throughput levels that are much higher than are possible in a standard ALOHA system.

VII. Conclusion

In this paper we have shown how an adaptive antenna array may be used to improve the performance of a slotted ALOHA packet radio network. The adaptive array creates a capture effect by separating packets in angle and thereby preventing collisions at the receiver. We described how an adaptive array could be used in a slotted system and analyzed the performance of such a system. Typical performance results were presented. It was shown that this technique achieves a performance level comparable to CSMA. Unlike CSMA, however, a slotted ALOHA system with an adaptive array does not require that all users be able to hear each other in order to attain high throughput. The performance is determined primarily by the array resolution, the number of nulls, and the length of the uncertainty interval in each slot.

Acknowledgment

Significant contributions to this work were made by M. Azizoğlu, Dr. F.D. Garber, G.M. Huffman, and Dr. H.C. Yu under a previous NASA contract [11] at The Ohio State University. The authors gratefully acknowledge their contributions.

References

[1] L. Kleinrock and S.S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.* vol. COM-23, pp. 410–422, Apr. 1975.
[2] A.B. Carleial and M.E. Hellman, "Bistable behavior of ALOHA-type systems," *IEEE Trans. Commun.*, vol. COM-23, pp. 401–410, Apr. 1975.
[3] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, pp. 117–128, Jan. 1977.
[4] L. Kleinrock and F.A. Tobagi, "Packet switching in radio channels: Part I—Carrier sense multiple access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400–1416, Dec. 1975.
[5] L.G. Roberts, "ALOHA packet system with and without slots and capture," *Comput. Commun. Rev.*, vol. 5, no. 2, pp. 199–204, Apr. 1975.
[6] C.C. Lee, "Random signal levels for channel access in packet broadcast networks," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1026–1034, July 1987.
[7] C. Namislo, "Analysis of mobile radio slotted ALOHA networks," *IEEE Trans. Vehic. Technol.*, vol. VT-33, pp. 199–204, Aug. 1984.
[8] R. Binder, S.D. Huffman, I. Gurantz, and P.A. Vena, "Crosslink architectures for a multiple satellite system," *Proc. IEEE*, vol. 75, pp. 74–82, Jan. 1987.
[9] R.T. Compton, Jr., *Adaptive Antennas—Concepts and Performance.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[10] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays.* New York: Wiley, 1980.
[11] M. Azizoğlu, R.T. Compton, Jr., F.D. Garber, G.M. Huffman, and H.C. Yu, "Adaptive arrays in satellite packet radio communication systems," Final Rep. 718163-1, The Ohio State University, Dep. Elec. Eng., ElectroSci. Lab., Nov. 1987.
[12] S.P. Applebaum, "Adaptive arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-24, pp. 585–598, Sept. 1976.
[13] B. Widrow, P.E. Mantey, L.J. Griffiths, and B.B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, pp. 2143–2159, Dec. 1967.
[14] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-10, pp. 853–862, Nov. 1974.
[15] S. Lin and D.J. Costello, *Error Control Coding: Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1983.
[16] M.I. Skolnik, *Radar Handbook.* New York: McGraw-Hill, 1970.
[17] S.W. Golomb, *Shift Register Sequences.* San Francisco, CA: Holden-Day, 1967.
[18] D.A. Davis and S.A. Gronemeyer, "Performance of slotted ALOHA random access with delay capture and randomized time of arrival," *IEEE Trans. Commun.*, vol. COM-28, pp. 703–710, May 1980.
[19] D.M. DiCarlo and R.T. Compton, Jr., "Reference loop phase shift in adaptive arrays," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-14, pp. 599–607, July 1978.
[20] D.M. DiCarlo, "Reference loop phase shift in an n-element adaptive array," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-15, pp. 576–582, July 1979.
[21] E. Çinlar, *Introduction to Stochastic Processes.* Englewood Cliffs, NJ: Prentice-Hall, 1975.
[22] D. Bertsekas and R.G. Gallager, *Data Networks.* Englewood Cliffs, NJ: Prentice-Hall, 1987.
[23] J.D.C. Little, "A proof for the queueing formula: $l = \lambda w$," *Oper. Res.*, vol. 9, pp. 383–387, May 1961.
[24] L. Kleinrock, *Queueing Systems, Volume II: Computer Applications.* NewYork: Wiley, 1976.

James Ward (S'84–M'91) was born in Belleville, NJ, on July 11, 1964. He received the B.E.E. degree from the University of Dayton, Dayton, OH in 1985, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the Ohio State University, Columbus, OH, in 1987 and 1990.

From 1985 to 1990 he worked as a Graduate Research Associate at the Ohio State University ElectroScience Laboratory. He is now a Member of the Technical Staff in the Radar Systems Group at M.I.T. Lincoln Laboratory. His current research interests include packet radio systems and communication and radar applications of adaptive antennas.

Dr. Ward is a member of Tau Beta Pi, Sigma Xi, and Phi Kappa Phi.

R.T. Compton, Jr. (S'58–M'59–SM'82–F'84) was born in St. Louis, MO, on July 26, 1935. He received the S.B. degree from Massachusetts Institute of Technology, Cambridge, in 1958, and the M.Sc. and Ph.D. degrees from the Ohio State University, Columbus, OH in 1961 and 1964, all in electrical engineering.

He is a Professor of Electrical Engineering at the Ohio State University. From 1965 to 1967, he was an Assistant Professor of Engineering at Case Institute of Technology, Cleveland, OH, and from 1967 to 1968 he was an NSF Postdoctoral Fellow at the Technische Hochschule, Munich, Germany. He spent the 1983-1984 academic year at the Naval Research Laboratory, Washington, DC.

Dr. Compton is a member of Sigma Xi and Pi Mu Epsilon. He received the M. Barry Carlton Award for best paper from the IEEE Aerospace and Electronic Systems Society in 1983.

A Study of Discretization Error in the Finite Element Approximation of Wave Solutions

Robert Lee, *Member, IEEE,* and Andreas C. Cangellaris, *Member, IEEE*

Abstract—**A dispersion analysis is used to study the errors caused by the spatial discretization of the finite element method for the two-dimensional scalar Helmholtz equation. It is shown that the error can be determined analytically for a uniform mesh of infinite extent. Numerical results are presented to show the effects of several parameters on the error. These parameters are the nodal density, the electrical size of mesh, the direction of propagation of the incident wave, the type of element, and the type of boundary condition.**

I. Introduction

In the past several years the use of the finite element method in electromagnetics has increased rapidly because of its versatility at handling very complex, arbitrary geometries. A primary consideration in any approximate numerical technique is the sources of potential errors in the solution. Without a good understanding of the causes of numerical error, one cannot have any confidence in the accuracy of their solution. In the finite element method, a major source of error is introduced by the spatial discretization of the problem domain into elements. Within each element, the behavior of the fields is described by a polynomial approximation. This approximation results in an error, which we will henceforth refer to as the discretization error.

Currently, there is a widespread belief that the solution accuracy is dependent mainly on the nodal density per wavelength if we exclude geometrical considerations. The nodal density used to generate results in the literature is usually between 10 and 20 nodes/λ where λ is the wavelength. Bayliss *et al.* [1] performed a mathematical and numerical study of discretization error in which they showed that the use of a fixed nodal density does not guarantee accuracy in the solution. They found that the accuracy also depends on the electrical size of the problem domain. Fang [2] did a similar numerical study for the finite-difference time-domain (FDTD) method and showed that the discretization error in FDTD has the same dependence as the finite element method on the electrical length of the problem domain. Recently, work has appeared on discretization error for the one-dimensional case [3], which confirms the results from [1].

In this paper, we seek to characterize the discretization error by means of a dispersion analysis for the two-dimensional problem. The dispersion analysis for many different elements has been done by several researchers [4]–[6]. From the dispersion analysis, we will show that the error can be obtained analytically for the case of an infinite mesh. Furthermore, we will demonstrate, either analytically or numerically, the effect of various parameters on the discretization error. Several factors, excluding geometrical considerations, which affect discretization error are the electrical size of the mesh, the electrical size of the elements in the mesh, the order of the function used to approximate the fields in each element, the boundary conditions which are applied, and the type of field excitation. In this paper, we will consider all these factors and explain their effects on the discretization error.

II. Analysis

Although this paper is primarily a numerical study of discretization error in the finite element method, a secondary goal is to analytically characterize the error in such a way that we can gain a better understanding of the causes and effects of the error. To this end, a simple geometry is chosen for the study. The geometry is a free space region on which a square mesh is embedded in the $x - y$ plane. The geometry is assumed to be two dimensional, so the fields do not vary in z. The excitation is a plane wave propagating in free space and passing through the mesh at an arbitrary angle ϕ^i in the $x - y$ plane. Because of the simplicity of this problem, the analytical solution is known, so comparisons can be easily made between the exact and the numerical solution.

The finite-element mesh covers an $a \times a$ square region. An example of such a mesh is shown in Fig. 1 where we use bilinear four-node square elements in the mesh. The nodal separation is h. For our numerical study, the number of elements in the mesh is variable. By selecting a uniform mesh, we remove the errors due to distorted elements. Furthermore, the analysis of the discretization error is greatly simplified. It should be noted that the numerical error is strongly dependent on the geometry under consideration. For example, finer discretization is necessary around a perfectly conducting corner because of the rapid field variation near the corner. We have purposely chosen our geometry to avoid

Manuscript received October 31, 1991; revised January 21, 1992. This work was supported in part by the Joint Services Electronics Program under Contract N00014-89-J-1007 with the Ohio State University Research Foundation.

R. Lee is with the ElectroScience Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, OH 93212.

A. C. Cangellaris is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721.

IEEE Log Number 9200442.

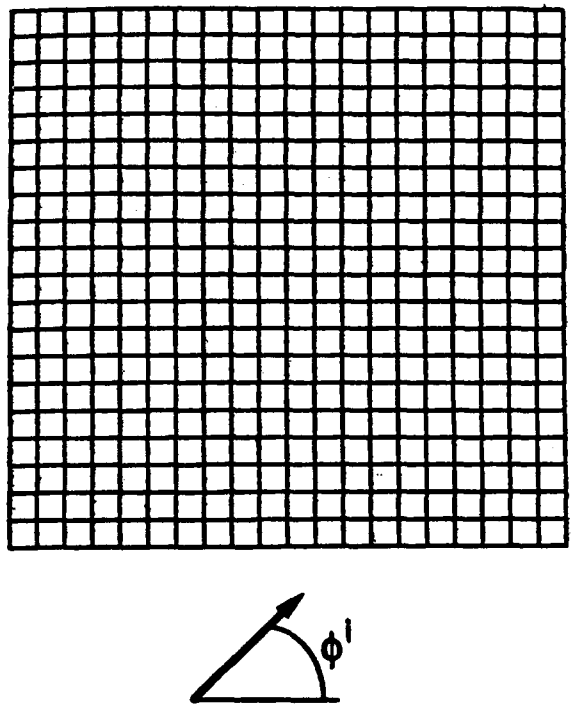


Fig. 1. An $a \times a$ finite element mesh square elements with sides of length h. The region encompassed by the mesh is assumed to be free space with a plane wave incident at an angle of ϕ^i.

such geometrical considerations. Instead, we seek to characterize discretization error over a region where the field variation is solely due to frequency. In this section, we will first briefly review the finite-element method. Then we will discuss the errors in terms of the dispersion relation associated with a given discretization.

A. Finite Element Formulation

For the geometry shown in Fig. 1, we must solve the two-dimensional Helmholtz equation. An application of the method of weighted residuals [7] guarantees that the Helmholtz equation is satisfied in the weak sense. Assuming a $TM_z(\vec{E} = \hat{x}E,\ \vec{H} = \hat{x}H_x,\ \hat{y}H_y)$ polarization, the weighted residual expression is given by

$$\iint \left[(\nabla_{xy}^2 + k^2)E\psi_j\right] dS = 0 \qquad (1)$$

where the weighting functions $\psi_j (j = 1, 2, \cdots)$ constitute a set of first-order differentiable, scalar weight functions which we choose. The variable k is the free space wavenumber, and $\nabla_{xy} = \hat{x}\partial/\partial x + \hat{y}\partial/\partial y$. By using Green's first identity, we get

$$\iint_S \left[\nabla_{xy}E \cdot \nabla_{xy}\psi_j - k^2 E\psi_j\right] dS - \int_{\partial S} \left[\psi_j(\hat{n} \cdot \nabla E)\right] dl = 0 \qquad (2)$$

where ∂S is the line enclosing S and $\hat{n}$ is the outward unit normal on ∂S. The field E_z can be represented by an appropriate set of expansion functions. Equation (2) is then applied to each element in the mesh where S is now the element itself and ∂S is the element boundary. Because the geometry is free space, the integral along ∂S at interelement boundaries cancels out during the assembly of element equations. Therefore, this term only exists on the outer boundary of the mesh and becomes the vehicle for the application of appropriate boundary conditions on the mesh boundary.

Using Galerkin's approximation, we choose the expansion functions to be the same as the weighting functions. We consider two specific elements, the isoparametric bilinear four-node quadrilateral and the isoparametric quadratic nine-node quadrilateral [7]. In order to simplify the analysis, the quadrilaterals are chosen to be squares. Both these elements lead to a matrix equation which is sparse and banded.

B. Dispersion Analysis

It is well known that finite-difference/finite element approximations to the wave equation are always dispersive. This means that the phase velocity of the numerical solution is frequency-dependent even in the absence of any dispersion in the actual media. We can obtain the numerical dispersion relation which relates the numerical wavenumber $\hat{k}$ to the exact free space wavenumber of the plane wave propagating through the grid. For the four-node square element, the derivation of the dispersion relation has been presented in [6]. We will review the analysis for the four-node square element and also consider a special case for the nine-node square element.

To remove the effects of the boundary condition from the dispersion analysis, we assume that the mesh is of infinite extent in x and y. We first consider the solution using the four-node bilinear square elements. For this choice of element, all the rows of the matrix equation are identical. Thus, the numerical dispersion relation is obtained by considering only the ith row in the matrix equation. The ith weighting function, which generates row i, has support only over the four elements that share node i, so let us consider the localized grid around node i shown in Fig. 2(a). The equation for row i is

$$\frac{(kh)^2}{36}\left[16E_i + \sum_{n=i+1}^{i+4} E_n + 4\sum_{n=i+5}^{i+8} E_n\right] + \frac{1}{6}\left[2\sum_{n=i+1}^{i+8} E_n - 16E_i\right] = 0 \qquad (3)$$

where E_n is the unknown field quantity at node n. Assuming a plane wave passing through the mesh at an angle ϕ^i with respect to the x axis, we can write our numerical solution at node n as

$$E_n(x_n, y_n) = e^{-j\hat{k}(x_n \cos\phi^i + y_n \sin\phi^i)} \qquad (4)$$

where (x_n, y_n) is the coordinate of node n. Substituting (4) into (3), we can solve for the numerical wavenumber $\hat{k}$. The general case requires the solution of a transcendental equation. We can consider two special cases. The first is the case where the plane wave is propagating along the x direction ($\phi^i = 0°$). We obtain the dispersion relation

$$\frac{\hat{k}}{k} = \frac{1}{kh}\cos^{-1}\left[\frac{1 - (kh)^2/3}{1 + (kh)^2/6}\right]. \qquad (5)$$

Because the expansion function for the fields varies linearly along the x direction, the above dispersion relation is equivalent to the one that we would obtain for the linear three-node triangular element. Also, since the fields associated with the plane wave only varies with x, (5) is the dispersion relation for the one-dimensional linear two-node element. The second case that we consider is the one where the plane wave is propagating diagonally across the square mesh ($\phi^i = 45°$). The dispersion relation for this case is

$$\frac{\hat{k}}{k} = \frac{\sqrt{2}}{kh}\cos^{-1}\left[\frac{1-(kh)^2/6}{1+(kh)^2/12}\right]. \tag{6}$$

From (5) and (6), the dependence of $\hat{k}$ on frequency is apparent. More specifically, $\hat{k}/k$ depends on kh where $\hat{k}/k \to 1$ as $kh \to 0$. In both cases, the ratio $\hat{k}/k \to 1$ as $kh \to 0$.

From Fig. 2(a), we see that node i is coupled to eight other nodes to form the resulting dispersion relations in (5) and (6). For the quadratic square element, the dispersion relation is significantly more complex since node i (Fig. 2(b)) is now coupled to 24 other nodes. It should be noted that h is the separation distance between the nodes rather than the width of the element. Unlike the bilinear element, the finite element approximation of the wave equation is different for different nodes in the quadratic element. The rows associated with nodes i to $i+8$ differ from those associated with nodes $i+9$ to $i+20$ which in turn differ from those associated with nodes $i+21$ to $i+24$. Thus, the dispersion analysis is not straightforward. To simplify the analysis, let us consider the special case of a plane wave traveling in the x direction. The dispersion relation in this case is the same as the one-dimensional dispersion relation for a three-node quadratic element (Fig. 3). As the figure depicts, the exterior nodes have been assigned odd numbers while the interior nodes have been assigned even numbers. The finite element equation for the odd node E_{2n+1} is

$$\begin{aligned}&-(5+2(kh)^2)(E_{2n-1}+E_{2n+3})\\&+(40+4(kh)^2)(E_{2n}+E_{2n+2})\\&+2(-35+8(kh)^2)E_{2n+1}=0\end{aligned} \tag{7}$$

while for the interior nodes E_{2n} and E_{2n+2} we obtain, respectively,

$$(10+(kh)^2)E_{2n-1}+(-20+8(kh)^2)E_{2n}+(10+(kh)^2E_{2n+1}=0 \tag{8}$$

$$(10+(kh)^2)E_{2n+1}+(-20+8(kh)^2)E_{2n+2}+(10+(kh)^2E_{2n+3}=0. \tag{9}$$

Equations (8) and (9) are then used in (7) to eliminate the interior nodes E_{2n}, E_{2n+2} and get

$$-\left[(kh)^4+4(kh)^2+15\right](E_{2n-1}+E_{2n+2})+2\left[3(kh)^4-26(kh)^2+15\right]E_{2n+1}=0. \tag{10}$$

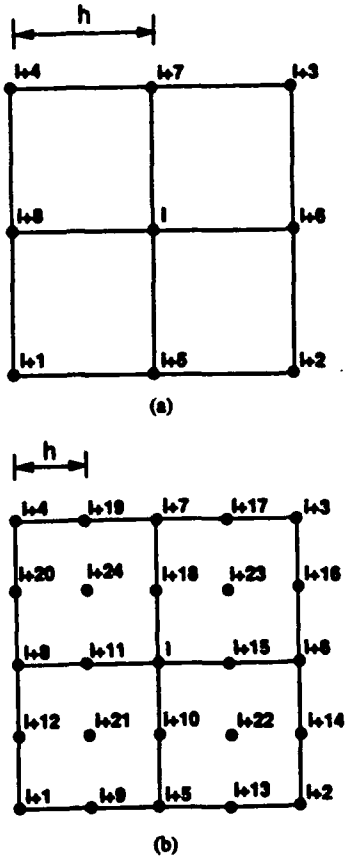


Fig. 2. A localized section of grids used in the dispersion analysis. The two elements considered are (a) the bilinear element and (b) the quadratic element.

Fig. 3. The grid used to evaluate the dispersion analysis for the one-dimensional quadratic element.

The above internal node condensation procedure allows us to obtain a reduced system of identical equations of the form (10) for all exterior (odd) nodes. A standard dispersion analysis of (10) produces the following dispersion relation:

$$\frac{\hat{k}}{k} = \frac{1}{kh}\cos^{-1}\left[\sqrt{\frac{2(kh)^4-11(kh)^2+15}{(kh)^4+4(kh)^2+15}}\right]. \tag{11}$$

A simple study of the expression under the radical reveals that for frequencies below cutoff ($kh \le \pi$) the numerical solution does not exhibit artificial damping, except for a cutoff region given by $\sqrt{2.5} \le kh \le \sqrt{3}$.

Without the introduction of boundary conditions, the error can now be determined from (5) and (11) for the case of a plane wave traveling along the x or y direction in a mesh with square elements which have either bilinear or quadratic variation. Furthermore, (6) can be used to find the error when the plane wave is traveling diagonal to the mesh composed of square elements with bilinear variation. The

exact solution E_z is given by

$$E_z = e^{-jk(x\cos\phi^i + y\sin\phi^i)}. \tag{12}$$

From our dispersion analysis, we find that the numerical solution $\hat{E}_z$ can be written as

$$\hat{E}_z = e^{-j\hat{k}(x\cos\phi^i + y\sin\phi^i)}. \tag{13}$$

The predicted error in the absence of boundary conditions does not simulate the actual error when the boundary conditions are present, but it provides an approximate quantitative measure of the expected error. Therefore, a chart that provides the error in the absence of boundary conditions can be very useful in choosing the nodal density for a given problem. From our analysis, a plane wave that travels a distance l, accumulates a phase error of $\theta_{err} = (k - \hat{k})l$ rad in the absence of boundary conditions. In Fig. 4, the normalized phase error (in degrees per wavelength) $\theta_{err}/\lambda = (k - \hat{k})$ is plotted as a function of the nodal density normalized by the wavelength, i.e., λ/h. Three curves are shown. Two of the curves correspond to an x propagating wave traveling in a mesh composed of either bilinear or quadratic square elements. The third curve is the case of a plane wave traveling diagonally along a mesh composed of bilinear square elements. From this figure, we can estimate the phase error for a given mesh in free space. As an example, consider a mesh similar to the one in Fig. 1, where $a = 10\lambda$. Assume that we are using bilinear square elements where the nodal separation is $\lambda/h = 10$. If a plane wave is incident at an angle of $\phi^i = 0°$, then we can estimate from Fig. 4 that there will be a phase error of 56° at the right border of the mesh relative to the left border of the mesh.

It should be noted that the finite element expansion function varies bilinearly along the diagonal direction, whereas the variation along the x and y direction is linear. Because of this variation, the largest phase error occurs at $\phi^i = n\,90°$, and the smallest error occurs at $\phi^i = 45° + n\,90°$ where n is an integer. The error for any other angle should be bounded by the error given for these two sets of angles. Thus, the numerical solution is dependent on the angle of incidence of the incoming wave. This numerical anisotropy has also been observed in the FDTD method [8]. It is also interesting to observe that the phase error is significantly less for the quadratic element than for the bilinear element. From the point of view of computational efficiency, we should be comparing both the phase error and the computation time associated with each element. In the next section, we perform these comparisons by running several numerical experiments.

III. Boundary Conditions

In the dispersion analysis for an infinite mesh, we showed that the discretization error can be strictly characterized as a phase error. For realistic problems of interest, the mesh must be truncated, and a boundary condition imposed at the mesh boundary. The boundary condition can significantly alter the discretization error that we obtained for the infinite mesh. To show this effect, we consider two boundary conditions. One is a radiation boundary condition that perfectly absorbs the plane wave. The second is the essential (Dirichlet) boundary condition, where we specify the value of the fields on the boundary. A third possibility is the natural (Neumann) boundary condition, but this case is not considered because the size of the error produced is exactly the same as the error from the essential boundary condition.

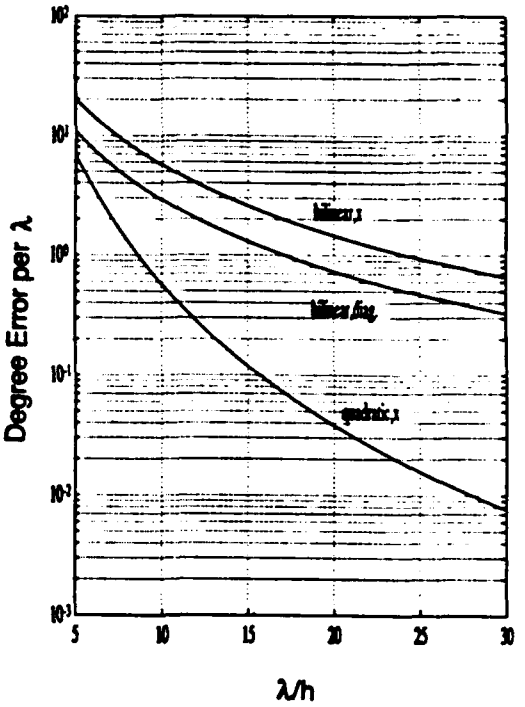


Fig. 4. Plot of the phase error for an infinite mesh. Results are shown for a plane wave propagating through the mesh at two different angles ($\phi^i = 0°$ and 45°) for the bilinear element and one angle for the quadratic element ($\phi^i = 0°$).

For the radiation boundary condition, we consider a plane wave propagating in both the diagonal and the x direction. The differential equation is given by

$$(\nabla_{xy}^2 + k^2)E(x, y) = 0 \tag{14}$$

where the boundary conditions for the diagonally propagating plane wave are

$$\begin{aligned} E(x = 0, y) &= e^{-jky/\sqrt{2}}, \\ \left(\frac{\partial}{\partial x} + j\frac{k}{\sqrt{2}}\right)E(x = a, y) &= 0, \\ E(x, y = 0) &= e^{jkx/\sqrt{2}}, \\ \left(\frac{\partial}{\partial y} + j\frac{k}{\sqrt{2}}\right)E(x, y = a) &= 0. \end{aligned} \tag{15}$$

For an x propagating wave, the boundary conditions are

$$\begin{aligned} &E(x = 0, y) = 1, \quad \left(\frac{\partial}{\partial x} + jk\right)E(x = a, y) = 0, \\ &\frac{\partial E(x, y = 0)}{\partial y} = 0, \qquad \frac{\partial E(x, y = a)}{\partial y} = 0. \end{aligned} \tag{16}$$

Before we calculate the numerical solutions to the above finite element problems, let us consider the case of the x propagating wave more closely. This problem is really one

dimensional. The related one-dimensional numerical problem for (14) and (16) is described by the differential equation

$$\left(\frac{d^2}{dx^2} + \hat{k}^2\right)\hat{E}(x) = 0 \tag{17}$$

with boundary conditions

$$\hat{E}(0) = 1 \quad \left(\frac{d}{dx} + jk\right)\hat{E}(a) = 0. \tag{18}$$

The term $\hat{E}$ is used to distinguish the numerical solution from the analytical one. The boundary conditions are the ones associated with the exact wavenumber k. Actually, $\hat{E}(x)$ does not satisfy (17), but rather the system of difference equations pertinent to the type of element used in the numerical discretization. Nevertheless, the model of (17) and (18) has a straightforward analytic solution that allows us to illustrate the effects of the radiation condition on the discretization error in a simple manner. The resulting solution is

$$\hat{E}_z = e^{-jkx} - j\frac{\delta}{\hat{k}}\sin \hat{k}x \tag{19}$$

where $\delta = k - \hat{k}$. δ can be assumed to be small if we consider nodal densities greater than five nodes/λ. From (19), we can see the effect of the radiation boundary condition directly. The first term in (19) is the solution for the case of an infinite mesh. The second term is the perturbation due to the boundary condition. The periodic behavior of the perturbation will be evident in the plots of error shown later in this paper. For the case of the diagonally propagating wave, there is no one-dimensional equivalent. Thus, we must rely on a numerical solution in this instance.

For the case where essential boundary conditions are applied, we consider the same differential equation as in (14) with the boundary conditions,

$$\begin{aligned} E(x = 0, y) &= e^{-jky\sin\phi^i}, \\ E(x = a, y) &= e^{-jk(a\cos\phi^i + y\sin\phi^i)} \\ E(x, y = 0) &= e^{-jkx\cos\phi^i}, \\ E(x, y = a) &= e^{-jk(x\cos\phi^i + a\sin\phi^i)}. \end{aligned} \tag{20}$$

Note that (20) describes the boundary conditions for a plane wave of arbitrary incidence. Unlike the radiation boundary condition, there is no situation where the numerical solution degenerates to the one-dimensional case. Although the analytical solution for the x propagating wave is one dimensional, the numerical errors due to the essential boundary condition are two dimensional; therefore, it is necessary to evaluate the solution numerically for the essential boundary conditions.

In order to describe the error, we introduce several error parameters. Let us define the exact solution to be $E = |E|e^{j\theta}$ and the numerical solution to be $\hat{E} = |\hat{E}|e^{j\hat{\theta}}$. Then we define the percentage magnitude error at some point (x_n, y_n) in the mesh to be

$$\text{magnitude error} = \left|\frac{|E(x_n, y_n)| - |\hat{E}(x_n, y_n)|}{|E(x_n, y_n)|}\right| \times 100\% \tag{21}$$

and the phase error at that same point to be

$$\text{phase error} = |\theta - \hat{\theta}|. \tag{22}$$

Another parameter of importance is the percentage rms error over the entire mesh, which we define to be

$$\text{rms error} = \frac{1}{a}\left[\int_0^a\int_0^a |E(x, y) - \hat{E}(x, y)|^2\, dx\, dy\right]^{1/2} \times 100\%. \tag{23}$$

To show the effects of various parameters on the discretization error, numerical solutions are generated for three square regions which differ from each other in electrical size ($a = 0.4\lambda$, 1.2λ, 4.0λ). Four nodal densities ($\lambda/h = 10$, 15, 20, and 25) are considered for a plane wave incident at angles of $\phi^i = 0°$ and $\phi^i = 45°$. Bilinear square elements are used to compute solutions for both boundary conditions. In addition, we generate numerical solutions for the essential boundary condition using quadratic square elements.

In Table I, we present numerical results for the radiation boundary condition using bilinear square elements. The columns marked Max. Mag., and Max. Phase indicate the maximum magnitude and phase error, respectively, over all the nodes in the mesh. The rms error is also shown in the table. The remaining two columns show the average magnitude and phase error over all the nodes. By examining the table it is evident that the majority of the error for $\phi^i = 0°$ is in the phase, as predicted by the one-dimensional analysis. In fact, the error is mostly phase error even for $\phi^i = 45°$. The amount of phase error is very close to the predictions from Fig. 4 for an infinite mesh. The size of the error is dependent on the electrical size of the mesh as well as the nodal density, which agrees with the results in [1]. It should also be noted that the error for a diagonally propagating wave is significantly less than for a wave propagating in the x direction. The distribution of the error across the mesh is not evident from the table. In Figs. 5 and 6, we plot the magnitude and phase errors as a function of position for the case $a = 1.2\ \lambda$, $\lambda/h = 20$, and $\phi^i = 0°$. We see that the phase error accumulates as the wave travels across the mesh. Also, a periodic ripple is present in both the magnitude and phase error plots as predicted from (19). The errors do not vary with y, so the numerical solution is one dimensional.

In Table II, we consider the numerical solution for the essential boundary condition using bilinear square elements. The size of the error as a function of nodal density, electrical size of mesh, and angle of incidence of the plane wave is similar to the case for the radiation boundary condition, but the characteristics of the error is significantly different. Both magnitude and phase error are present. In addition, the growth of the error with increasing mesh size is considerably faster than for the radiation boundary condition. The distribution of the magnitude and phase error over the mesh are shown in Figs. 7 and 8. The geometrical parameters (a, λ/h, ϕ^i) are the same as those shown in Figs. 5 and 6. The

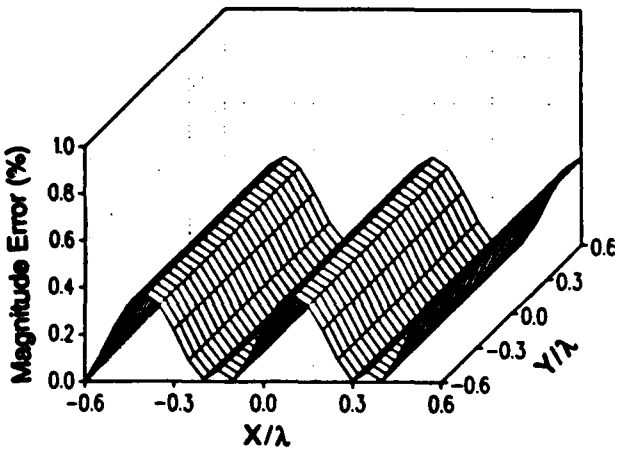


Fig. 5. Plot of the magnitude error of E_z over the grid for the radiation boundary condition.

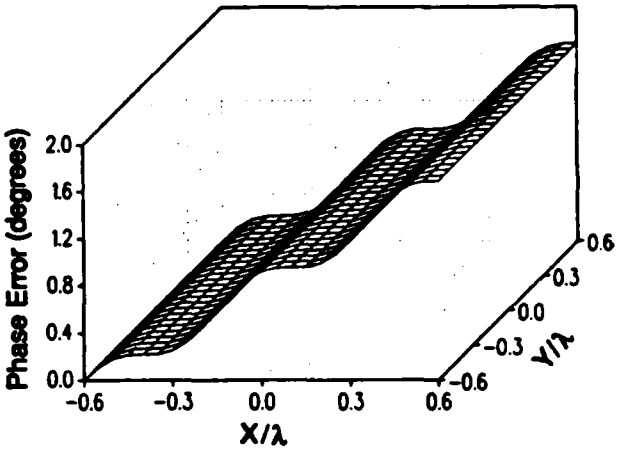


Fig. 6. Plot of the phase error of E_z over the grid for the radiation boundary condition.

TABLE I
ERRORS ASSOCIATED WITH THE RADIATION BOUNDARY CONDITION FOR BILINEAR ELEMENTS

a	λ/h	φ' (deg)	Max. Mag. (%)	Average Mag. (%)	Max. phase (deg)	Average Phase (deg)	RMS Error (%)
	10	0	0.90	0.50	2.74	1.60	3.42
0.4 λ	15	0	0.46	0.24	1.23	0.73	1.54
	20	0	0.27	0.14	0.70	0.42	0.87
	25	0	0.17	0.09	0.45	0.26	0.56
	10	45	1.32	0.40	0.95	0.41	1.10
0.4 λ	15	45	0.58	0.18	0.43	0.19	0.50
	20	45	0.32	0.10	0.25	0.11	0.28
	25	45	0.21	0.06	0.16	0.07	0.18
	10	0	1.38	0.67	6.46	3.11	6.62
1.2 λ	15	0	0.64	0.32	2.97	1.44	3.03
	20	0	0.37	0.18	1.69	0.82	1.72
	25	0	0.24	0.12	1.09	0.53	1.11
	10	45	2.02	0.57	3.88	1.43	3.21
1.2 λ	15	45	0.90	0.25	1.75	0.65	1.45
	20	45	0.51	0.14	0.99	0.37	0.82
	25	45	0.32	0.09	0.64	0.24	0.53
	10	0	1.46	0.68	23.07	11.70	23.58
4.0 λ	15	0	0.72	0.35	10.42	5.24	10.63
	20	0	1.12	0.24	5.91	2.97	5.98
	25	0	0.26	0.13	3.78	1.89	3.83
	10	45	3.39	0.83	14.76	5.35	11.40
4.0 λ	15	45	1.51	0.37	6.64	2.41	5.17
	20	45	0.85	0.21	3.75	1.37	2.92
	25	45	0.55	0.13	2.41	0.88	1.88

TABLE II
ERRORS ASSOCIATED WITH THE ESSENTIAL BOUNDARY CONDITION FOR BILINEAR ELEMENTS

a	λ/h	φ' (deg)	Max. Mag. (%)	Average Mag. (%)	Max. Phase (deg)	Average Phase (deg)	RMS Error (%)
	10	0	2.06	0.51	0.32	0.06	0.89
0.4 λ	15	0	0.91	0.27	0.14	0.04	0.42
	20	0	0.51	0.17	0.08	0.03	0.25
	25	0	0.33	0.11	0.05	0.02	0.16
	10	45	1.03	0.25	0.15	0.03	0.44
0.4 λ	15	45	0.46	0.14	0.07	0.02	0.21
	20	45	0.26	0.09	0.04	0.01	0.12
	25	45	0.16	0.06	0.03	0.01	0.08
	10	0	29.37	7.11	13.00	3.56	14.42
1.2 λ	15	0	12.50	2.92	5.86	1.66	6.09
	20	0	6.78	1.65	3.31	0.95	3.38
	25	0	4.25	1.06	2.12	0.62	2.16
	10	45	12.40	2.83	8.06	1.58	6.17
1.2 λ	15	45	5.03	1.25	3.37	0.72	2.64
	20	45	2.79	0.71	1.91	0.41	1.48
	25	45	1.77	0.46	1.21	0.27	0.94
	10	0	75.82	24.30	49.78	17.40	45.39
4.0 λ	15	0	62.27	19.11	28.49	10.63	32.86
	20	0	46.20	14.40	18.02	6.85	23.94
	25	0	34.32	10.89	12.27	4.84	17.72
	10	45	69.65	15.00	32.14	6.60	27.17
4.0 λ	15	45	23.20	5.30	13.06	2.70	9.96
	20	45	12.06	2.81	7.21	1.53	5.35
	25	45	7.54	1.79	4.80	0.99	3.36

errors are in the form of large standing waves in both the magnitude and phase with a one wavelength periodicity along the direction of propagation. We see that the errors are two dimensional, which produces a two-dimensional numerical solution to a one-dimensional problem. The differences in the numerical results between the two boundary conditions imply that the effect of the boundary condition on discretization error is variable; therefore, in choosing a boundary truncation scheme, we must consider not only the accuracy and numerical efficiency of the scheme but also its effect on the discretization error. For $a = 4\lambda$, we see from Table II that the essential boundary conditions produce unacceptable errors even for a nodal density of 25 nodes/λ where the rms error is 17.7% for $\phi' = 0$. The corresponding error for the radiation boundary condition is only 3.8%. It is evident that the solution with the radiation boundary condition is more computationally efficient for the simple geometry considered here.

The next consideration is the effect of changing the type of element used in the mesh. The errors are presented in Table III for the quadratic nine-node square element with essential boundary conditions. A similar table is not considered for the radiation boundary condition because the errors in this case are basically the same as those obtained from the infinite mesh geometry. If we compare the errors obtained in Table II to those shown in Table III, we observe one major

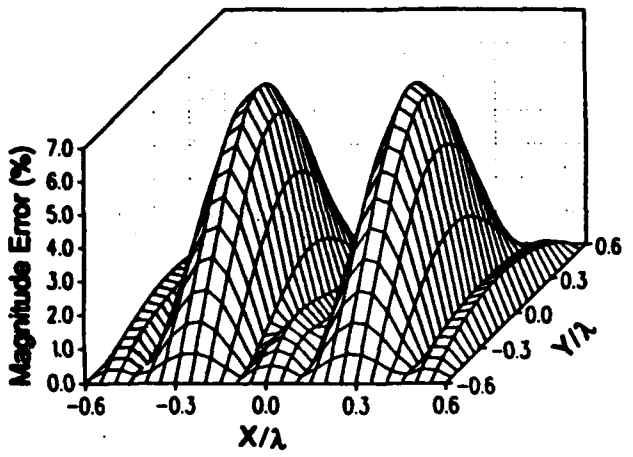


Fig. 7. Plot of the magnitude error of E_z over the grid for the essential boundary condition.

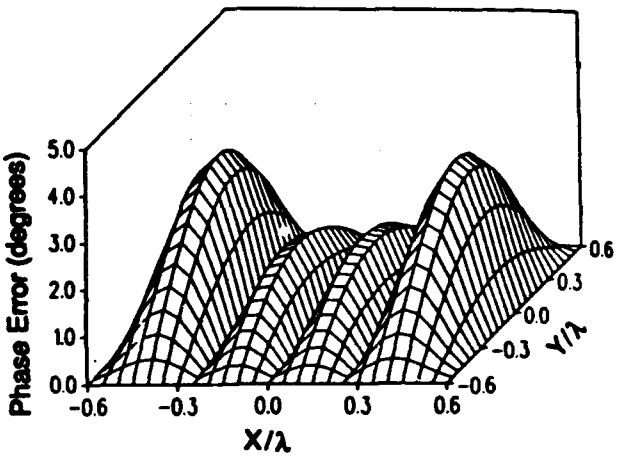


Fig. 8. Plot of the phase error of E_z over the grid for the essential boundary condition.

difference. The size of the error has been substantially reduced by the use of quadratic elements. In many instances, the error has been reduced by a factor of ten.

As we mentioned earlier in this paper, a major concern in the finite element method is the computational time necessary to obtain a solution for a given accuracy. Usually, most of the computation time is spent in the solution of the finite element matrix equation. Assuming the use of a banded matrix solver, the number of floating point operations needed to solve the matrix equation is dependent on the number of unknowns and the half-bandwidth of the matrix. The numerical errors in Table II and III are based on a nodal density rather than element density, so the number of unknowns is the same for both types of elements. The difference is in the half-bandwidth of the matrix. From Fig. 2, we see that for the bilinear element, each node is connected to eight other nodes while for the quadratic element, each node connects to 24 other nodes. Thus, the half-bandwidth of the matrix for the quadratic elements is greater than the half-bandwidth of the matrix for the bilinear element. In Table IV, we record the CPU time (for a Cray YMP) needed to fill, factor, and solve the finite element matrix equation for the cases where we use bilinear and quadratic square elements with essential boundary conditions. Also, the table provides the rms error and the half-bandwidth of the corresponding matrix equation. From the table, we see that the half-bandwidth increases by a factor of two when the elements are switched from bilinear to quadratic. The corresponding increase in the computation time is anywhere from a factor of 1.03 to a factor of three. Although the computation time is greater for the quadratic element, the gain in the accuracy of the numerical solution suggests that this element is much more computationally efficient. As an example, consider the case where $a = 4\lambda$. The solution obtained for quadratic elements with a nodal density of $\lambda/h = 10$ is considerably better than the solution obtained for bilinear elements with $\lambda/h = 25$ while the computation time for the bilinear elements is seven times greater than the quadratic case. Another consideration is the memory requirements for a given problem. Although the half-bandwidth is larger for quadratic elements, the corresponding increase in accuracy means that a smaller nodal density is necessary for a given accuracy requirement. In most cases, the increase in memory requirements due to the larger bandwidth for quadratic elements is more than outweighted by a decrease in memory requirements due to the smaller number of unknowns.

TABLE III
ERRORS ASSOCIATED WITH THE ESSENTIAL BOUNDARY CONDITION FOR QUADRATIC ELEMENTS

a	λ/h	ϕ^i (deg)	Max. Mag. (%)	Average Mag. (%)	Max. Phase (deg)	Average Phase (deg)	RMS Error (%)
	10	0	0.180	0.054	0.026	0.005	0.092
0.4 λ	15	0	0.051	0.014	0.005	0.001	0.022
	20	0	0.016	0.005	0.001	4.2e-4	0.007
	25	0	0.007	0.002	5.8e-4	1.8e-4	0.003
	10	45	0.060	0.018	0.008	0.002	0.031
0.4 λ	15	45	0.018	0.005	0.001	5.0e-4	0.007
	20	45	0.006	0.002	5.2e-4	1.6e-4	0.008
	25	45	0.002	7.2e-4	2.2e-4	7.2e-5	0.003
	10	0	2.300	0.560	1.250	0.330	1.183
1.2 λ	15	0	0.490	0.120	0.260	0.072	0.247
	20	0	0.155	0.039	0.082	0.024	0.080
	25	0	0.066	0.016	0.034	0.010	0.033
	10	45	0.560	0.130	0.380	0.074	0.283
1.2 λ	15	45	0.110	0.028	0.075	0.016	0.058
	20	45	0.036	0.009	0.024	0.005	0.019
	25	45	0.015	0.004	0.010	0.002	0.008
	10	0	21.30	7.36	7.00	2.95	11.83
4.0 λ	15	0	5.70	1.84	1.68	0.69	2.92
	20	0	1.91	0.62	0.54	0.22	0.97
	25	0	0.79	0.26	0.23	0.09	0.40
	10	45	2.25	0.52	1.36	0.30	1.05
4.0 λ	15	45	0.48	0.11	0.27	0.06	0.21
	20	45	0.15	0.04	0.09	0.02	0.07
	25	45	0.06	0.01	0.04	0.01	0.03

IV. CONCLUSION

From the results of this study on discretization error, there are several observations to make. First, the fact that the discretization error increases with the electrical size of the mesh has serious implications with respect to the approximate radiation boundary conditions [9], which require the mesh to

TABLE IV
COMPARISON OF ERRORS BETWEEN THE BILINEAR AND QUADRATIC CASE. THE COMPUTATATION TIMES NECESSARY TO FILL, FACTOR, AND SOLVE THE CORRESPONDING FINITE ELEMENT MATRICES ARE ALSO LISTED

			Bilinear			Quadratic		
a	λ/h	ϕ^i (deg)	HBW	CPU (s)	RMS Error (%)	HBW	CPU (s)	rms Error (%)
	10	0	9	5.9e-3	0.89	14	6.1e-3	0.092
0.4 λ	15	0	13	1.4e-2	0.42	22	1.6e-2	0.022
	20	0	17	2.6e-2	0.25	30	3.1e-2	0.007
	25	0	21	4.3e-2	0.16	38	5.3e-2	0.003
	10	45	9	5.9e-3	0.44	14	6.1e-3	0.031
0.4 λ	15	45	13	1.4e-2	0.21	22	1.6e-2	0.007
	20	45	17	2.6e-2	0.12	30	3.1e-2	0.008
	25	45	21	4.3e-2	0.08	38	5.3e-2	0.003
	10	0	25	6.5e-2	14.42	46	8.5e-2	1.183
1.2 λ	15	0	37	0.172	6.09	70	0.26	0.247
	20	0	49	0.383	3.38	94	0.70	0.080
	25	0	61	0.712	2.16	118	1.44	0.033
	10	45	25	6.5e-2	6.17	46	8.5e-2	0.283
1.2 λ	15	45	37	0.172	2.64	70	0.26	0.058
	20	45	49	0.383	1.48	94	0.70	0.019
	25	45	61	0.712	0.94	118	1.44	0.008
	10	0	81	1.68	45.39	158	3.82	11.83
4.0 λ	15	0	121	5.29	32.86	238	13.3	2.92
	20	0	161	16.10	23.94	318	47.20	0.97
	25	0	201	29.00	17.72	398	86.30	0.40
	10	45	81	1.68	27.17	158	3.82	1.05
4.0 λ	15	45	121	5.29	9.96	238	13.30	0.21
	20	45	161	16.10	5.35	318	47.20	0.07
	25	45	201	29.00	3.36	398	86.30	0.03

extend a significant distance away from the geometry of interest in order to obtain an accurate solution. More rigorous truncation techniques, such as the hybrid FEM/integral equation technique [10] or the bymoment method [11], allows for the use of a coarser mesh to obtain the same accuracy because the mesh can be truncated close to the geometry of interest. The second observation is that the solution with quadratic elements is always more computationally efficient than the solution with bilinear elements. Even the memory storage requirements are usually smaller for quadratic elements. A third point to note is the effect of the boundary condition on discretization error. The results indicate that radiation boundary conditions produce lower levels of error than essential boundary conditions. Finally, we show that the discretization error can be analytically characterized as phase error for an infinite mesh. The curves in Fig. 4 can be used as an initial estimate on the size of the error as a function of nodal density, element type, and the electrical size of the mesh with the understanding that errors due to geometry and boundary conditions are not incorporated into the curves.

ACKNOWLEDGMENT

Computer time on the Cray YMP was provided by the Ohio Supercomputer Center.

REFERENCES

[1] A. Bayliss, C. I. Goldstein, and E. Turkel, "On accuracy conditions for the numerical computation of waves," NASA ICASE Rep. 84-38, Aug. 1984.
[2] J. Fang, "Time domain finite difference computations for Maxwell's equations," Ph.D. dissertation, Dept. Elec. Eng. Comput. Sci., Univ. California, Berkeley, CA, 1989.
[3] A. F. Peterson and R. J. Baca, "Error in the finite element discretization of the scalar Helmholtz equation over electrically large regions," *IEEE Microwave Guided Wave Lett.*, vol. 1, pp. 219-222, Aug. 1991.
[4] G. W. Platzman, "Some response characteristics of finite element tidal models," *J. Comp. Phys.*, vol. 40, pp. 36-63, 1981.
[5] R. Mullen and T. Belytschko, "Dispersion analysis of finite element semidiscretization of the two-dimensional wave equation," *Int. J. Num. Methods Eng.*, vol. 18, pp. 11-29, 1982.
[6] D. R. Lynch, K. D. Paulsen, and J. W. Strohbehn, "Finite element solution of Maxwell's equation for hyperthemia treatment planning," *J. Comput. Phys.*, vol. 58, pp. 246-269, 1985.
[7] E. B. Becker, G. F. Carey, and J. T. Oden, *Finite Elements: An Introduction*, vol. 1. Englewood Cliffs, NJ: Prentice-Hall, 1981.
[8] A. C. Cangellaris, "Time-domain computation of electromagnetic wave scattering by the method of conforming boundary elements," Ph.D. dissertation, Dept. Elec. Eng. Comput. Sci., Univ. California, Berkeley, CA, 1985.
[9] A. Bayliss, M. Gunzberger, and E. Turkel, "Boundary conditions for the numerical solution of elliptic equations in exterior regions," *SIAM J. Appl. Math.*, vol. 42, pp. 430-451, 1982.
[10] B. H. McDonald and A. Wexler, "Finite element solution of unbounded field problems," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, pp. 841-847, 1972.
[11] A. C. Cangellaris and R. Lee, "The bymoment method for two-dimensional electromagnetic scattering," *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 1429-1437, 1990.
[12] R. Lee, "Rigorous grid truncation for the finite element solution of electromagnetic scattering problems," Ph.D. dissertation, Dept. Elec. Comput. Eng., Univ. Arizona, Tucson, AZ, 1990.

Robert Lee (S'82-M'84-S'85-M'90), for a photograph and biography please see page 1437 of the September 1990 issue of this TRANSACTIONS.

Andreas C. Cangellaris (M'86), for a photograph and biography please see page 1437 of the September 1990 issue of this TRANSACTIONS.

Two-Dimensional Angle and Polarization Estimation Using the ESPRIT Algorithm

Jian Li and R. T. Compton, Jr.

Abstract—**It is shown how the ESPRIT algorithm may be used with a square array of crossed dipoles to estimate both the two-dimensional arrival angles and the polarizations of incoming narrow-band signals. The ESPRIT algorithm exploits the invariance properties of such an array so that both angle and polarization estimates may be computed. Some typical examples showing the use of this approach are presented.**

I. Introduction

The ESPRIT algorithm [1] can be used with an array of cross-polarized elements to estimate both the directions of arrival (DOA) and the polarizations of incoming plane waves. In a previous paper [2], the authors described how to use ESPRIT with a uniform linear array of crossed dipoles to compute signal directions in one spatial angle, along with the polarizations. In this communication, we generalize these results to two spatial angles. The arriving signals are assumed to be narrow-band, so the signals received on different array elements differ only by a phase factor.

In this correspondence, we consider a square array consisting of L^2 pairs of crossed short dipoles. We show how the ESPRIT algorithm can be used with such an array to estimate signal directions in two spatial angles and signal polarizations. We also illustrate the effects of signal direction and polarization on the performance of the estimator.

II. Problem Formulation

We consider the array shown in Fig. 1. The array consists of L^2 pairs of crossed dipoles, or a total of $2L^2$ elements. The signal from each dipole is to be processed separately. The ilth dipole pair, where $i, l = 1, 2, \cdots, L$, has its center at $(x, y) = ((i-1)\delta, (l-1)\delta)$. The distance δ between adjacent dipoles is assumed to be a half-wavelength. At the ilth dipole pair, $x_{il}(t)$ denotes the signal received on the x-axis dipole and $y_{il}(t)$ the signal received on the y-axis dipole.

Suppose K (with $K \le L^2$) continuous wave (CW) signals impinge on the array from angular directions (θ_k, ϕ_k), $k = 1, 2, \cdots, K$, where r, θ, and ϕ denote standard polar coordinates, as shown in Fig. 1. Assume each signal has an arbitrary elliptical electromagnetic polarization [5].

To specify the signal polarizations, we use the following definitions. Given a transverse electromagnetic (TEM) wave propagating into the array, we consider the polarization ellipse produced by the electric field as the incoming wave is viewed from the coordinate origin. Note that unit vectors $\mathbf{e}_\phi, \mathbf{e}_\theta, -\mathbf{e}_r$, in that order, form a right-handed coordinate system for an incoming signal. Suppose the electric field has transverse components E_ϕ and E_θ:

$$\mathbf{E} = E_\phi \mathbf{e}_\phi + E_\theta \mathbf{e}_\theta. \tag{1}$$

(We call E_ϕ the horizontal component and E_θ the vertical component of the field.) In general, as time progresses, E_ϕ and E_θ describe a polarization ellipse as shown in Fig. 2. Given this ellipse, we define α and β to be the ellipticity and the orientation angles, respectively (see [2]). We define β to be in the range $0 \le \beta < \pi$. α is always in the range $-\pi/4 \le \alpha \le \pi/4$.

For a given signal polarization, specified by α and β, the electric field components are given by (aside from a common phase factor)

$$E_\phi = E\cos\gamma, \quad E_\theta = E\sin\gamma e^{j\eta} \tag{2}$$

where γ and η can be computed from α and β [6], [2]. γ is always in the range $0 \le \gamma \le \pi/2$, and η is in the range $-\pi \le \eta < \pi$. α and β can also be computed from γ and η [6], [2].

Thus an arbitrary plane wave coming into the array may be characterized by four angular parameters and an amplitude. For example, the kth signal, $k = 1, 2, \cdots, K$, will be characterized by its arrival angles (θ_k, ϕ_k), its polarization ellipticity angle α_k and orientation angle β_k, and its amplitude E_k (i.e., E_k is the value of E in (2) for the kth signal). We will say the kth signal is defined by $(\theta_k, \phi_k, \alpha_k, \beta_k, E_k)$.

We assume that each dipole in the array is a short dipole whose output voltage is proportional to the electric field along the dipole. An incoming signal with arbitrary electric field components E_ϕ and

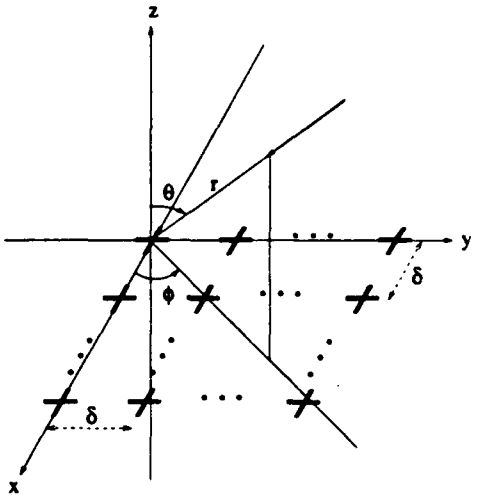


Fig. 1. Crossed dipoles.

Manuscript received December 17, 1990; revised January 31, 1992.

This work was supported in part by the Joint Services Electronics Program under Contract N00014-89-J-1007 with The Ohio State University Research Foundation.

Jian Li was with the Department of Electrical Engineering, The Ohio State University, Columbus, OH. She is now with the Department of Electrical Engineering, University of Kentucky, Lexington, KY 40506.

R. T. Compton, Jr. is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210.

IEEE Log Number 9200359.

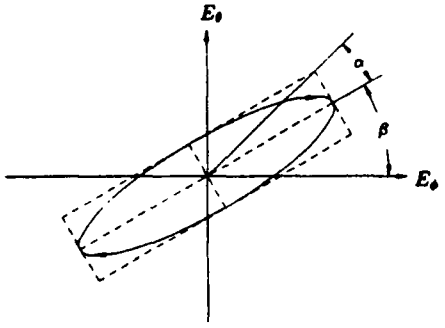


Fig. 2. Polarization ellipse.

E_θ has x, y, z components:

$$\begin{aligned}\mathbf{E} &= E_\phi \mathbf{e}_\phi + E_\theta \mathbf{e}_\theta \\ &= \left(E_\theta \cos\theta \cos\phi - E_\phi \sin\phi\right)\mathbf{e}_x \\ &\quad + \left(E_\theta \cos\theta \sin\phi + E_\phi \cos\phi\right)\mathbf{e}_y \\ &\quad + \left(E_\theta \sin\theta\right)e_z. \end{aligned} \tag{3}$$

When E_ϕ and E_θ are expressed in terms of E, γ, and η as in (2), the electric field components become

$$\begin{aligned}\mathbf{E} = E\big[&\left(\sin\gamma \cos\theta \cos\phi e^{j\eta} - \cos\gamma \sin\phi\right)\mathbf{e}_x \\ &+ \left(\sin\gamma \cos\theta \sin\phi e^{j\eta} + \cos\gamma \cos\phi\right)\mathbf{e}_y \\ &- \left(\sin\gamma \sin\theta e^{j\eta}\right)\mathbf{e}_z\big]. \end{aligned} \tag{4}$$

Let us define the space phase factors

$$p = e^{j\frac{2\pi\delta}{\lambda}\sin\theta\cos\phi}, \quad q = e^{j\frac{2\pi\delta}{\lambda}\sin\theta\sin\phi}, \tag{5}$$

where λ is the wavelength of the signal. Including the time and space phase factors in (4), we find that an incoming signal characterized by $(\theta, \phi, \alpha, \beta, E)$ produces a signal vector in the crossed dipole pair centered at $(x, y) = ((i-1)\delta, (l-1)\delta)$ as follows:

$$\mathbf{z}_{il} = \begin{bmatrix} x_{il}(t) \\ y_{il}(t) \end{bmatrix} = \mathbf{u}s(t)p^{i-1}q^{l-1} \tag{6}$$

where

$$\mathbf{u} = \begin{bmatrix} \sin\gamma \cos\theta \cos\phi e^{j\eta} - \cos\gamma \sin\phi \\ \sin\gamma \cos\theta \sin\phi e^{j\eta} + \cos\gamma \cos\phi \end{bmatrix} \tag{7}$$

and

$$s(t) = Ee^{j(\omega t + \psi)} \tag{8}$$

with ω the frequency and ψ the carrier phase of the signal at the coordinate origin at $t = 0$.

We assume that K such signals, specified by $(\theta_k, \phi_k, \alpha_k, \beta_k, E_k)$, $k = 1, 2, \cdots, K$, are incident on the array. The carrier phase angles ψ_k are assumed to be random variables, each uniformly distributed on $[0, 2\pi)$ and all statistically independent of each other. In addition we assume a thermal noise voltage vector $\mathbf{n}_{il}(t)$ is present in each signal vector $\mathbf{z}_{il}(t)$. The $\mathbf{n}_{il}(t)$ are assumed to be zero mean, complex Gaussian processes statistically independent of each other, with covariance $\sigma^2\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix.

Under these assumptions, the total signal vector received by the crossed dipole pair centered at $(x, y) = ((i-1)\delta, (l-1)\delta)$ is given by

$$\mathbf{z}_{il}(t) = \sum_{k=1}^{K} \mathbf{u}_k s_k(t) p_k^{i-1} q_k^{l-1} + \mathbf{n}_{il}(t), \qquad i, l = 1, 2, \cdots, L \tag{9}$$

where $\mathbf{u}_k$, p_k, q_k, and $s_k(t)$ are given by (7), (5), (5), and (8), respectively, with subscript k added to the amplitude E and to each angular quantity.

Let $\mathbf{z}(t)$, $\mathbf{s}(t)$, and $\mathbf{n}(t)$ be column vectors containing the received signals, incident signals, and noise, respectively, i.e

$$\mathbf{z}(t) = \left[\mathbf{z}_{11}^T(t) \ \cdots \ \mathbf{z}_{1L}^T(t) \ \cdots \ \mathbf{z}_{L1}^T(t) \ \cdots \ \mathbf{z}_{LL}^T(t)\right]^T \tag{10}$$

$$\mathbf{s}(t) = \left[s_1(t) \ s_2(t) \ \cdots \ s_K(t)\right]^T \tag{11}$$

$$\mathbf{n}(t) = \left[\mathbf{n}_{11}^T(t) \ \cdots \ \mathbf{n}_{1L}^T(t) \ \cdots \ \mathbf{n}_{L1}^T(t) \ \cdots \ \mathbf{n}_{LL}^T(t)\right]^T \tag{12}$$

where $(\cdot)^T$ denotes the transpose. The received signal vector is then

$$\mathbf{z}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{13}$$

where $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_K]$ is a $2L^2 \times K$ matrix with each column given by

$$\begin{aligned}\mathbf{a}_k = \big[&\mathbf{u}_k^T \ \mathbf{u}_k^T q_k \ \cdots \ \mathbf{u}_k^T q_k^{L-1} \ \mathbf{u}_k^T p_k \ \mathbf{u}_k^T p_k q_k \\ &\cdots \ \mathbf{u}_k^T p_k q_k^{L-1} \ \cdots \ \cdots \ \mathbf{u}_k^T p_k^{L-1} \ \mathbf{u}_k^T p_k^{L-1} q_k \\ &\cdots \ \mathbf{u}_k^T p_k^{L-1} q_k^{L-1}\big]^T. \end{aligned} \tag{14}$$

The columns $\mathbf{a}_k$ are assumed linearly independent. They define a K-dimensional *signal subspace* in a $2L^2$-dimensional space.

By assuming linearly independent columns in $\mathbf{A}$, we are excluding from consideration degenerate cases, such as when two signals arrive from the same direction with the same polarization, when more than two signals of arbitrary polarization arrive from the same direction, or when a signal yields zero output at the x- and y-axis dipoles at the same time.

We assume that the element signals are sampled at N distinct times t_n, $n = 1, 2, \cdots, N$. The problem of interest is to determine the quantities $(\theta_k, \phi_k, \alpha_k, \beta_k)$, $k = 1, 2, \cdots, K$, from the measurements $\mathbf{z}(t_n)$, $n = 1, 2, \cdots, N$.

III. Application of the ESPRIT Algorithm

The array geometry described above possesses several invariance properties that may be exploited by the ESPRIT algorithm. We shall first consider the case where the array covariance matrix and the number of incident signals are known. We then consider the practical situation where only a finite number of data samples is available and the number of incident signals is unknown.

The array covariance matrix has the form

$$\mathbf{R} = E\{\mathbf{z}(t)\mathbf{z}^H(t)\} = \mathbf{R}_0 + \sigma^2\mathbf{I}, \tag{15}$$

where

$$\mathbf{R}_0 = \mathbf{A}\mathbf{R}_s\mathbf{A}^H. \tag{16}$$

$(\cdot)^H$ denotes the complex conjugate transpose, $\mathbf{R}_s = E\{\mathbf{s}(t)\mathbf{s}^H(t)\}$ is the source covariance matrix, and $E(\cdot)$ denotes expectation.

From the array covariance matrix, the signal directions and polarizations may be calculated as follows [1], [7]. Let $\lambda_1 + \sigma^2 \geq \lambda_2 + \sigma^2 \geq \cdots \geq \lambda_K + \sigma^2 > \sigma^2 = \cdots = \sigma^2$ be the eigenvalues of $\mathbf{R}$, and $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_K, \mathbf{e}_{K+1}, \cdots, \mathbf{e}_{2L^2}$ be the corresponding orthonormal eigenvectors. Since the noise contribution to $\mathbf{R}$ for this ideal case is simply $\sigma^2\mathbf{I}$, the eigenvectors of $\mathbf{R}$ are also the

eigenvectors of $\mathbf{R}_0$. It can be shown that the columns in $\mathbf{E}_s = [\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_K]$ span the same signal subspace as the column vectors in $\mathbf{A}$ [3]. Therefore, there exists a unique nonsingular $\mathbf{T}$ such that

$$\mathbf{E}_s = \mathbf{AT}. \tag{17}$$

The columns in $\mathbf{E}_s$ are the *signal subspace eigenvectors*. The signal directions and polarizations are computed from them.

Consider first the calculation of the space factors p_k, $k = 1, 2, \cdots, K$. From Fig. 1, we note that the overlapping subarrays consisting of the first and the last $L-1$ y-axis rows of the crossed dipoles are the same except for the displacement δ parallel to the x-axis. For the kth incident signal, the displacement δ results in the space factor p_k. The subvectors of $\mathbf{a}_k$ consisting of the first and the last $2L(L-1)$ elements of $\mathbf{a}_k$ differ by the factor p_k, as can also be seen from (14). (Note that $\mathbf{u}_k$ is a two-element column vector, as in (7).) Let $\mathbf{A}_{p1}$ and $\mathbf{A}_{p2}$ be the $2L(L-1) \times K$ submatrices of $\mathbf{A}$ consisting of the first and the last $2L(L-1)$ rows of $\mathbf{A}$, respectively. Then $\mathbf{A}_{p2} = \mathbf{A}_{p1}\Phi_p$, where Φ_p is the diagonal matrix with diagonal elements p_k, $k = 1, 2, \cdots, K$. Let $\mathbf{E}_{p1}$ and $\mathbf{E}_{p2}$ be the $2L(L-1) \times K$ submatrices formed from $\mathbf{E}_s$ in the same way that $\mathbf{A}_{p1}$ and $\mathbf{A}_{p2}$ are formed from $\mathbf{A}$. Then the diagonal elements p_k, $k = 1, 2, \cdots, K$, of Φ_p are the eigenvalues of the unique matrix $\Psi_p = \mathbf{T}^{-1}\Phi_p\mathbf{T}$ that satisfies [1]

$$\mathbf{E}_{p2} = \mathbf{E}_{p1}\Psi_p. \tag{18}$$

Consider next the calculation of the space factors q_k, $k = 1, 2, \cdots, K$. From Fig. 1, we note that the overlapping subarrays consisting of the left and the right $L-1$ x-axis columns of the crossed dipoles are the same except for a displacement δ parallel to the y-axis. For the kth incident signal, the displacement δ results in the space factor q_k, as can also be seen from (14). The subvectors of $\mathbf{a}_k$ consisting of the elements of $\mathbf{a}_k$ numbered $2L(i-1)+l$ and $2L(i-1)+l+2$, respectively, for $l = 1, 2, \cdots, 2(L-1)$ and $i = 1, 2, \cdots, L$, differ by the factor q_k. Let $\mathbf{A}_{q1}$ and $\mathbf{A}_{q2}$ be the $2L(L-1) \times K$ submatrices of $\mathbf{A}$ consisting of the rows of $\mathbf{A}$ numbered $2L(i-1)+l$ and $2L(i-1)+l+2$, respectively, $l = 1, 2, \cdots, 2(L-1) i = 1, 2, \cdots, L$. Then $\mathbf{A}_{q2} = \mathbf{A}_{q1}\Phi_q$, where Φ_q is the diagonal matrix with diagonal elements q_k, $k = 1, 2, \cdots, K$. Let $\mathbf{E}_{q1}$ and $\mathbf{E}_{q2}$ be the $2L(L-1) \times K$ submatrices formed from $\mathbf{E}_s$ in the same way that $\mathbf{A}_{q1}$ and $\mathbf{A}_{q2}$ are formed from $\mathbf{A}$. Then the diagonal elements q_k, $k = 1, 2, \cdots, K$, of Φ_q are the eigenvalues of the unique matrix $\Psi_q = \mathbf{T}^{-1}\Phi_q\mathbf{T}$ that satisfies

$$\mathbf{E}_{q2} = \mathbf{E}_{q1}\Psi_q. \tag{19}$$

Finally, consider the calculation of the ratios r_k from which the polarization angles may be calculated, where r_k is the ratio between the first and the second elements of $\mathbf{u}_k$, i.e.,

$$r_k = \frac{\sin\lambda_k \cos\theta_k \cos\phi_k e^{j\eta_k} - \cos\gamma_k \sin\phi_k}{\sin\gamma_k \cos\theta_k \sin\phi_k e^{j\eta_k} + \cos\gamma_k \cos\phi_k}, \quad k = 1, 2, \cdots, K. \tag{20}$$

From Fig. 1, we note that both dipoles in a given crossed dipole pair have the same space factors p_k or q_k. Moreover, in any dipole pair, the y-axis dipole output is related to the x-axis dipole output by a factor r_k, as seen in (6) and (7). Because of this, the subvectors of $\mathbf{a}_k$ consisting of the even and the odd numbered elements of $\mathbf{a}_k$ differ by the factor r_k. Let $\mathbf{A}_{r1}$ and $\mathbf{A}_{r2}$ be the $L^2 \times K$ submatrices of $\mathbf{A}$ consisting of the even and the odd numbered rows of $\mathbf{A}$, respectively. Then $\mathbf{A}_{r2} = \mathbf{A}_{r1}\Phi_r$, where Φ_r is the diagonal matrix with diagonal elements r_k, $k = 1, 2, \cdots, K$. Let $\mathbf{E}_{r1}$ and $\mathbf{E}_{r2}$ be the $L^2 \times K$ submatrices formed from $\mathbf{E}_s$ in the same way that $\mathbf{A}_{r1}$ and $\mathbf{A}_{r2}$ are formed from $\mathbf{A}$. Then the diagonal elements r_k, $k = 1, 2, \cdots, K$, of Φ_r are the eigenvalues of the unique matrix $\Psi_r = \mathbf{T}^{-1}\Phi_r\mathbf{T}$ that satisfies

$$\mathbf{E}_{r2} = \mathbf{E}_{r1}\Psi_r. \tag{21}$$

With multiple incident signals, it is necessary to determine the grouping of the eigenvalues of Ψ_p, Ψ_q, and Ψ_r, i.e., we must determine which eigenvalues of Ψ_p correspond to which eigenvalues of Ψ_q and to which eigenvalues of Ψ_r. One way could be to use the MUSIC spectrum [3] to determine the grouping [8]. This method, however, requires finding the extra eigenvectors $\mathbf{e}_{K+1}$, $\mathbf{e}_{K+2}, \cdots, \mathbf{e}_{2L^2}$ and searching over K^3 possibilities. Instead, we introduce a different approach below.

The proper grouping of the eigenvalues of Ψ_p, Ψ_q, and Ψ_r may be determined in two steps: 1) by pairing the eigenvalues of Ψ_p, and Ψ_q, and 2) by pairing the eigenvalues of Ψ_p and Ψ_r. To pair the eigenvalues of Ψ_p and Ψ_q, note that

$$\Psi_{pq} = \Psi_q^{-1}\Psi_p = \mathbf{T}^{-1}\Phi_q^{-1}\Phi_p\mathbf{T}. \tag{22}$$

Thus the eigenvalues $\bar{q}_k$ of Ψ_{pq}, $k = 1, 2, \cdots, K$, are the ratios between the eigenvalues of Ψ_p and their corresponding eigenvalues of Ψ_q. Therefore, for $k_1 = 1, 2, \cdots, K$, the eigenvalue of Ψ_q that corresponds to the eigenvalue p_{k_1} of Ψ_p is the element in the set $\{q_{k_2}, k_2 = 1, 2, \cdots, K\}$ that corresponds to the minimum of

$$\left\{\left|\frac{p_{k_1}}{q_{k_2}} - \bar{q}_{k_3}\right|, k_2, k_3 = 1, 2, \cdots, K\right\}. \tag{23}$$

Similarly, the eigenvalues of $\Psi_{pr} = \Psi_r^{-1}\Psi_p$, $\bar{r}_k$, $k = 1, , \cdots, K$, are also the ratios between the eigenvalues of Ψ_p and their corresponding eigenvalues of Ψ_r. Thus the eigenvalues of Ψ_p and Ψ_r can be paired in the same way as those of Ψ_p and Ψ_q. From the paired sets of eigenvalues of Ψ_p and Ψ_q and of Ψ_p and Ψ_r, we can determine the grouping (p_k, q_k, r_k), $k = 1, 2, \cdots, K$.

The arrival angles, ellipticity angles, and orientation angles can be computed from the sets (p_k, q_k, r_k), $k = 1, 2, \cdots, K$. The arrival angles (θ_k, ϕ_k), $k = 1, 2, \cdots, K$, are calculated from p_k and q_k as

$$\theta_k = \sin^{-1}\left\{\frac{\lambda}{2\pi\delta}\left[\arg^2(p_k) + \arg^2(q_k)\right]^{1/2}\right\} \tag{24}$$

$$\phi_k = \tan^{-1}\left[\frac{\arg(q_k)}{\arg(p_k)}\right]. \tag{25}$$

To determine the ellipticity angles α_k and the orientation angles β_k, we must first find γ_k and η_k from r_k. $\gamma_k \in [0, \pi/2]$ and $\eta_k \in [-\pi, \pi)$ can be determined from

$$\gamma_k = \tan^{-1}(|\xi_k|), \tag{26}$$

$$\eta_k = \arg(\xi_k), \tag{27}$$

where

$$\xi_k = \frac{r_k \cos\phi_k + \sin\phi_k}{\cos\theta_k(-r_k \sin\phi_k + \cos\phi_k)}. \tag{28}$$

From γ_k and η_k, we can determine $\alpha_k \in [-\pi/4, \pi/4]$ and $\beta_k \in [0, \pi)$ [6], [2].

This approach may be used to estimate the signal directions and polarizations as long as the matrices $\mathbf{A}$, $\mathbf{A}_{p1}$, $\mathbf{A}_{q1}$, and $\mathbf{A}_{r1}$ are all of full column rank, a condition that is satisfied in most cases. We shall restrict our consideration to the cases where this approach is applicable. For these cases, when the array covariance matrix is known exactly, the signal directions and polarizations can be found *exactly*.

In practical situations, however, only a finite number of noisy measurements are made at the dipole outputs. The estimates of the signal directions and polarizations must then be made from the available measurements. Also, the number of incident signals is unknown and must be estimated from the measurements. We shall use the minimum description length (MDL) criterion described by Wax and Kailath [9] to estimate the number of incident signals. The total least squares (TLS) algorithm [1], [10] is then used to estimate $\hat{\Psi}_p$, $\hat{\Psi}_q$, and $\hat{\Psi}_r$. The estimation steps in the signal subspace approach are as follows.

1) Compute the maximum likelihood estimate of the array covariance matrix $\mathbf{R}$

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{z}(t_n)\mathbf{z}^H(t_n) \tag{29}$$

where N denotes the number of measurements.
2) Compute the eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{2L^2}$ of $\hat{\mathbf{R}}$.
3) Estimate the number of incident signals $\hat{K}$ using the MDL criterion.
4) Obtain $\hat{\mathbf{E}}_s$ whose columns are the eigenvectors of $\hat{\mathbf{R}}$ that correspond to the $\hat{K}$ largest eigenvalues of $\hat{\mathbf{R}}$.
5) Form $\hat{\mathbf{E}}_{p1}$, $\hat{\mathbf{E}}_{p2}$, $\hat{\mathbf{E}}_{q1}$, $\hat{\mathbf{E}}_{q2}$, $\hat{\mathbf{E}}_{r1}$, and $\hat{\mathbf{E}}_{r2}$ from $\hat{\mathbf{E}}_s$ in the same way that $\mathbf{E}_{p1}$, $\mathbf{E}_{p2}$, $\mathbf{E}_{q1}$, $\mathbf{E}_{q2}$, $\mathbf{E}_{r1}$, and $\mathbf{E}_{r2}$ are formed from $\mathbf{E}_s$.
6) Calculate the TLS solution $\hat{\Psi}_p$ [1], [10] from $\hat{\mathbf{E}}_{p1}$ and $\hat{\mathbf{E}}_{p2}$.
7) Calculate the TLS solution $\hat{\Psi}_q$ from $\hat{\mathbf{E}}_{q1}$ and $\hat{\mathbf{E}}_{q2}$.
8) Calculate the TLS solution $\hat{\Psi}_r$ from $\hat{\mathbf{E}}_{r1}$ and $\hat{\mathbf{E}}_{r2}$.
9) Compute $\hat{p}_{k_1}$, $\hat{q}_{k_2}$, $\hat{\bar{q}}_{k_3}$, $\hat{r}_{k_4}$, and $\hat{\bar{r}}_{k_5}$, $k_1, k_2, k_3, k_4, k_5 = 1, 2, \cdots, \hat{K}$, by determining the eigenvalues of $\hat{\Psi}_p$, $\hat{\Psi}_q$, $\hat{\Psi}_{pq} = \hat{\Psi}_q^{-1}\hat{\Psi}_p$, $\hat{\Psi}_r$, and $\hat{\Psi}_{pr} = \hat{\Psi}_r^{-1}\hat{\Psi}_p$, respectively.
10) Determine the grouping scheme $(\hat{p}_k, \hat{q}_k, \hat{r}_k)$, $k = 1, 2, \cdots, \hat{K}$, as discussed in (22)-(23).
11) Calculate the direction and polarization estimates $(\hat{\theta}_k, \hat{\phi}_k, \hat{\alpha}_k, \hat{\beta}_k)$, $k = 1, 2 \cdots, \hat{K}$, from $(\hat{p}_k, \hat{q}_k, \hat{r}_k)$ as discussed in (24)-(28).

IV. Simulation Results

We show below several examples of the use of this technique for direction and polarization estimation. These results were obtained by using 50 Monte Carlo simulations. All incident signals are assumed to have the same unit amplitude E_k. The signal-to-noise ratio (SNR) used in the simulations is $-10\log_{10}\sigma^2 = 20$ dB. The number of data samples taken at each dipole output was $N = 31$.

Before presenting these simulation examples, however, we first describe the method we shall use to describe the accuracy of the estimates. For polarization estimates, we define the estimation error to be the spherical distance between the two points M and $\hat{M}$ on the Poincaré sphere that represent the actual signal polarization (α, β) and the estimated polarization $(\hat{\alpha}, \hat{\beta})$ [6], [2]. Similarly, we define the error ζ_d of a direction estimate $(\hat{\theta}, \hat{\phi})$ to be the angular separation between $(\hat{\theta}, \hat{\phi})$ and the actual arrival angle (θ, ϕ) as measured at the coordinate origin in Fig. 1. The direction and polarization estimate variances plotted below are the mean-squared values of these errors.

There is one other feature that must be handled. For the special case where $\theta = 0$, the angle ϕ has the same effect on the signal as the angle β. When $\theta = 0$, changing either ϕ or β simply rotates the principal axis of the polarization ellipse. (The principal axis is defined relative to $\mathbf{e}_\phi$.) To eliminate this ambiguity between ϕ and β, we always define $\beta = 0$ when $\theta = 0$.

We now present a few examples. We begin with the case of a single incident signal. The array used for this case consists of four

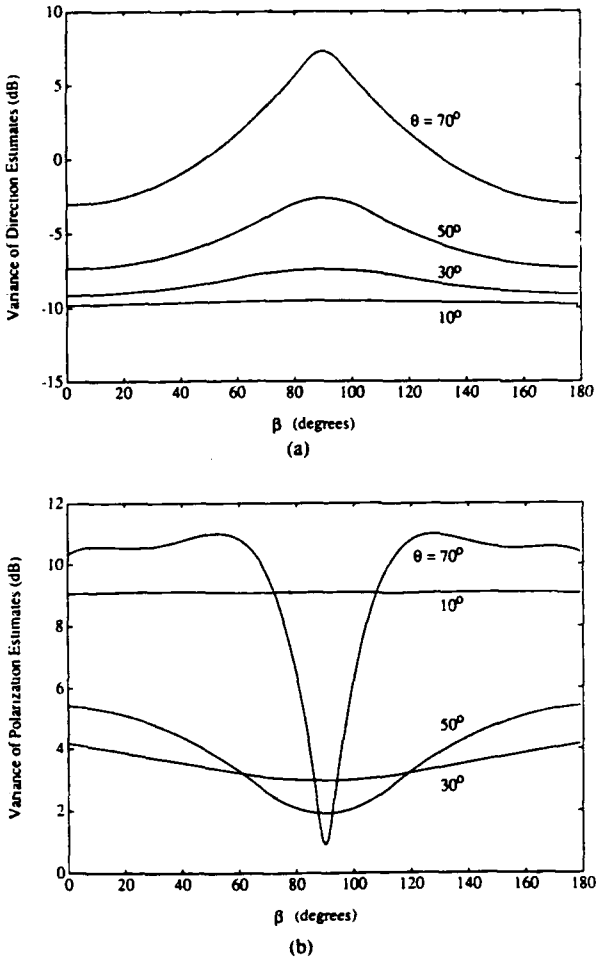


Fig. 3. Variance of estimates versus β for a linearly polarized signal ($\alpha = 0°$, $\phi = 0°$, SNR = 20 dB, N = 31). (a) Variance of direction estimates. (b) Variance of polarization estimates.

($L = 2$) pairs of crossed dipoles. Consider first a linearly polarized signal ($\alpha = 0°$) that arrives from azimuth angle $\phi = 0°$ and elevation angle θ. (For the case of one incident signal, we drop the subscript k.)

For such a signal, Fig. 3(a) shows the variance (in decibels with respect to degree squared) of the direction estimates as a function of β for several elevation angles θ. (The curves shown are the averages of the results for the 50 Monte Carlo simulations.) Note first that β has little effect on the direction estimates when θ is small. The reason is that for small θ, the outputs of the x- and y-axis dipoles are not close to zero at the same time. For large θ, however, β has more effect. When θ is large, the total power received by the x- and y-axis dipoles becomes small as β approaches 90°. This causes the performance of the direction estimates to deteriorate. Also, note from Fig. 3(a) that the accuracy of the direction estimates becomes worse as θ increases. This result occurs because the estimates $\hat{\theta}$ become worse due to the arcsin $(\cdot)$ in (24). As $\sin^{-1}(w)$ approaches 90°, the slope of $\sin^{-1}(w)$ approaches infinity. When θ is near 90°, a small perturbation in the argument of the arcsin in (24) causes a large error in $\hat{\theta}$.

Fig. 3(b) shows the variance (in decibels with respect to degree squared) of the polarization estimates as a function of β for several values of θ. In general, if we start with θ near zero and then increase θ, the accuracy of the polarization estimates at first improves with θ but then finally becomes worse as θ nears 90°. The

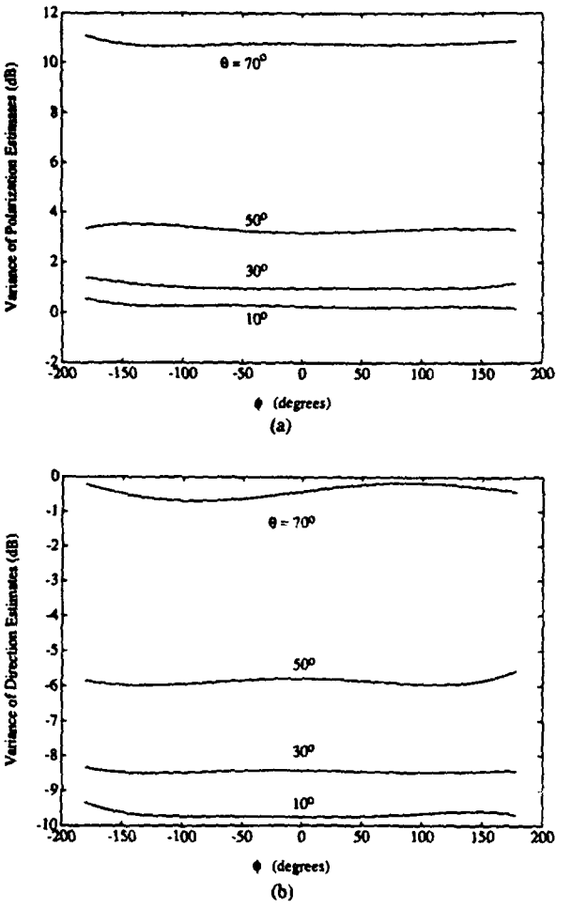


Fig. 4. Variance of estimates versus ϕ for a circularly polarized signal ($\alpha = 45°$, SNR = 20 dB, N = 31). (a) Variance of direction estimates. (b) Variance of polarization estimates.

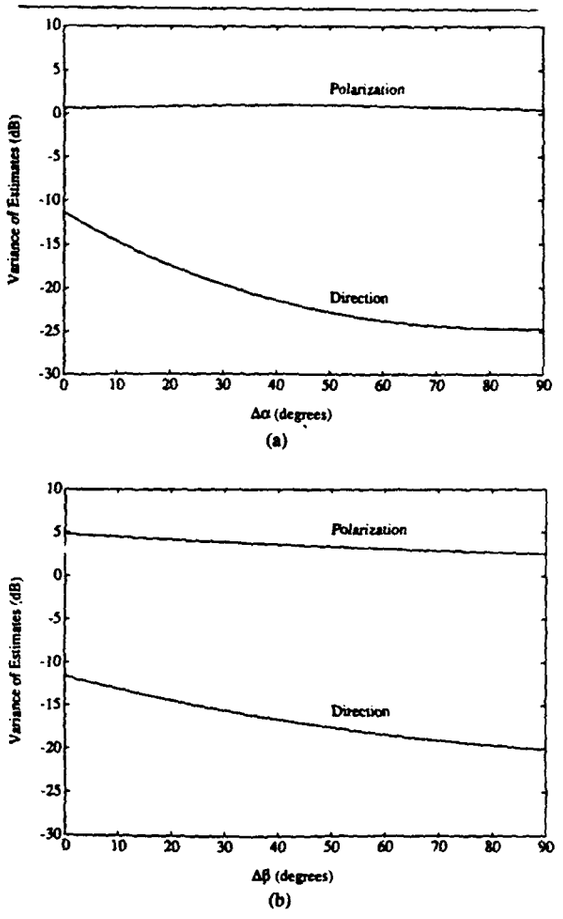


Fig. 5. Variance of estimates for the first of the two signals from ($\theta_1 = 5°$, $\phi_1 = 0°$) and ($\theta_2 = 10°$, $\phi_2 = 0°$) with SNR = 20 dB and $N = 31$. Variance versus (a) $\Delta\alpha$ when ($\alpha_1 = 45°$, $\beta_1 = 0°$) and ($\alpha_2 = 45° - \Delta\alpha$, $\beta_2 = 0°$), (b) $\Delta\beta$ when ($\alpha_1 = 30°$, $\beta_1 = 0°$) and ($\alpha_2 = 30°$, $\beta_2 = 0° + \Delta\beta$).

improvement with θ continues up to a higher value of θ when β is near 90° than for other values of β. This behavior is a result of the complicated dependence of the polarization estimates on $\hat{\theta}$, $\hat{\phi}$, and $\hat{\gamma}$, as seen from (28).

Fig. 4 shows another example, for a single circularly polarized signal ($\alpha = 45°$). Fig. 4 shows the variance of the direction and polarization estimates as a function of ϕ for several θ. Note that the accuracy of both the direction and polarization estimates depends little on ϕ but is better for small θ.

Now we present examples that illustrate how the separation in direction and polarization between two incident signals affect the performance of the estimator. The array used in the examples consists of 25 ($L = 5$) pairs of crossed dipoles.

Consider first the case where the two incident signals have closely spaced arrival angles with $\theta_1 = 5°$ and $\theta_2 = 10°$ and $\phi_1 = \phi_2 = 0°$. Fig. 5(a) shows the variances of the direction and polarization estimates of the first signal as functions of $\Delta\alpha$ when $\alpha_1 = 45°$ and $\alpha_2 = 45° - \Delta\alpha$ and $\beta_1 = \beta_2 = 0°$. Fig. 5(b) shows the variances of the direction and polarization estimates of the first signal as functions of $\Delta\beta$ when $\alpha_1 = \alpha_2 = 30°$ and $\beta_1 = 0°$ and $\beta_2 = 0° + \Delta\beta$.

Fig. 5 illustrates that the direction estimates are improved significantly by increasing the polarization separation. The reason is that when the two signals arrive from closely spaced directions with similar polarizations, the columns of matrix $\mathbf{A}$ in (13) become almost identical. $\mathbf{A}$, $\mathbf{A}_{\rho 1}$, $\mathbf{A}_{q1}$, and $\mathbf{A}_{\gamma 1}$ then become ill-conditioned. This ill-conditioning makes the signal subspace approach more sensitive to noise. For example, an ill-conditioned $\mathbf{A}$ results in an ill-conditioned $\mathbf{R}_0$ in (16). Since the noise contribution to $\hat{\mathbf{R}}$ is different from $\sigma^2\mathbf{I}$ when the number of data samples is finite, the signal subspace eigenvectors of $\hat{\mathbf{R}}$ are perturbed by the noise from the true eigenvectors of $\mathbf{R}_0$. The ill-conditioning of $\mathbf{R}_0$ then makes the signal subspace eigenvectors of $\mathbf{R}_0$ more sensitive to this perturbation [11].

On the other hand, Fig. 5 shows that the polarization estimates improve very little with larger polarization separation. The reason is that the ill-conditioning of $\mathbf{A}_{\gamma 1}$ is not reduced by increasing the polarization separation.

Next consider a case of two circularly polarized incident signals with $\alpha_1 = \alpha_2 = 45°$. Fig. 6(a) shows the variances of the direction and polarization estimates of the first signal as functions of $\Delta\theta$ when the signals arrive from elevation angles $\theta_1 = 5°$ and $\theta_2 = 5° + \Delta\theta$ and azimuth angles $\phi_1 = \phi_2 = 0°$. Fig. 6(b) shows the variances of the direction and polarization estimates of the first signal as functions of $\Delta\phi$ when the signals arrive from elevation angles $\theta_1 = \theta_2 = 30°$ and azimuth angles $\phi_1 = 0°$ and $\phi_2 = 0° + \Delta\phi$.

Fig. 6 shows that both the polarization estimates and the direction

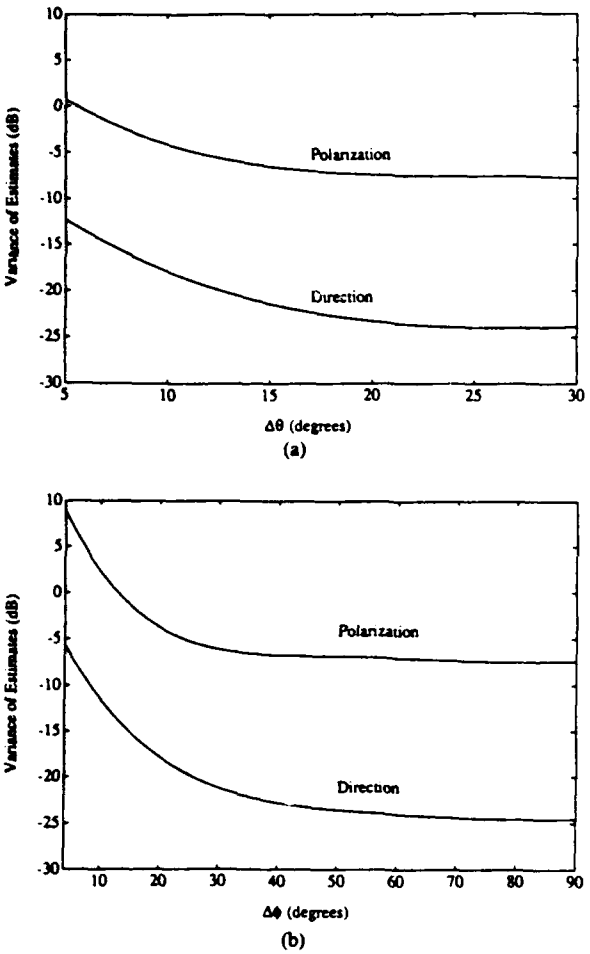


Fig. 6. Variance of estimates for the first of the two circularly signals with $\alpha_1 = \alpha_2 = 45°$, SNR = 20 dB, and $N = 31$. Variance versus (a) $\Delta\theta$ when ($\theta_1 = 5°$, $\phi_1 = 0°$) and ($\theta_2 = 5° + \Delta\theta$, $\phi_2 = 0°$), (b) $\Delta\phi$ when ($\theta_1 = 30°$, $\phi_1 = 0°$) and ($\theta_2 = 30°$, $\phi_2 = 0° + \Delta\phi$).

estimates have been improved by increasing the direction separation. Increasing the direction separation of the two signals reduces the ill-conditioning of $\mathbf{A}$, $\mathbf{A}_{p1}$, $\mathbf{A}_{q1}$, and $\mathbf{A}_{r1}$.

We remark that the MDL criterion provided accurate estimates of the number of incident signals K for all simulations conducted.

V. Conclusion

We have described the use of the ESPRIT algorithm for estimating two-dimensional arrival angles and polarizations of arbitrarily polarized signals with a square array of crossed dipoles. The ESPRIT algorithm exploits the invariance properties of such an array so that both angle and polarization estimates may be computed. Some typical examples showing the use of this approach have been presented.

Acknowledgment

The authors gratefully acknowledge several helpful discussions with Dr. Jim Ward.

References

[1] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech Signal Processing*," vol. 37, pp. 984-995, July 1989.

[2] J. Li and R. T. Compton, Jr., "Angle and polarization estimation using ESPRIT with a polarization sensitive array," *IEEE Trans. Antennas Propagat.*, vol. 39, pp. 1376-1383, Sept. 1991.

[3] "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276-280, Mar. 1986.

[4] P. Stoica and A. Nehorai, "Performance comparison of subspace rotation and MUSIC methods for direction estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 446-453, Feb. 1991.

[5] C. A. Balanis, *Antenna Theory—Analysis and Design*. New York: Harper & Row, 1982.

[6] G. A. Deschamps, "Geometrical representation of the polarization of a plane electromagnetic wave," *Proc. IRE*, vol. 39, pp. 540-544, May 1951.

[7] S. U. Pillai, *Array Signal Processing*. New York: Springer-Verlag, 1989.

[8] M. D. Zoltowski and D. Stavrinides, "Sensor array signal processing via a procrustes rotations based eigenanalysis of the ESPRIT data pencil," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 832-861, June 1989.

[9] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387-392, Apr. 1985.

[10] G. H. Golub and C. F. V. Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, pp. 883-893, Dec. 1980.

[11] G. W. Stewart, *Introduction to Matrix Computations*. New York: Academic, 1973.

Uniform GTD solution for the diffraction by metallic tapes on panelled compact-range reflectors

G.A. Somers
P.H. Pathak

Indexing terms: Panelled reflectors, Compact range, Diffraction, Scattering

Abstract: Metallic tape is commonly used to cover the interpanel gaps which occur in panelled compact-range reflectors. It is therefore of interest to study the effect of the scattering by the tape on the field in the target zone of the range. An analytical solution is presented for the target zone fields scattered by 2D metallic tapes. It is formulated by the generalised scattering matrix technique in conjunction with the Wiener–Hopf procedure. An extension to treat 3D tapes can be accomplished using the 2D solution via the equivalent current concept. The analytical solution is compared with a reference moment method solution to confirm the accuracy of the former.

1 Introduction

In present-day compact-range systems the main reflector may be physically very large for operation over a wide bandwidth, necessitating the reflector to be manufactured in sections. Once these panelled sections are aligned, one can use metallic tape to cover the interpanel gaps, as shown in Figs. 1 and 2. It is therefore of interest to study the effect of the scattering by the tape on the fields in the target zone of the range. Previously, Gupta and Burnside [1] have examined this problem in two dimensions (2D) by the method of moments (MoM); however, this is a low-frequency technique and therefore there are practical limitations due to computational effort on the electrical size of the reflector. As the frequencies of interest rise, or the reflectors become electrically large, the MoM becomes inefficient. In three dimensions the MoM solution proves to be intractable due to the large electrical area of typical compact-range reflectors. This scattering problem is well suited for analysis by a high-frequency technique such as the geometrical theory of diffraction (GTD) or one of its uniform versions. Within the framework of the GTD, scattered fields from a three-dimensional (3D) geometry can easily be derived from a 2D diffraction coefficient, providing that the radius of curvature of the reflector is large compared to a wavelength. This can be accomplished using the equivalent current method [2–4] or the incremental diffraction coefficients [5, 6]. Gupta and Burnside [1] empirically constructed a 2D diffraction coefficient for the fields scattered by a metallic tape by examining the trends of the data from various moment method cases.

Paper 8796H (E11(P)), first received 4th November 1991 and in revised form 17th February 1992

The authors are with the ElectroScience Laboratory, The Ohio State University, 1320 Kinnear Road, Columbus, OH 43212, USA

In this paper an asymptotically rigorous uniform diffraction coefficient for the tape geometry will be developed for the 2D case by superimposing the fields scattered by the two edges of the tape (steps). The fields diffracted from each step are found by combining the fields scattered from two simpler canonical geometries via the generalised scattering matrix technique (GSMT) [7, pp. 207–210]. This method produces an analytical solution which is computationally efficient, and, furthermore, the computation time is independent of the electrical size of the structure, making this an ideal solution technique for the electrically large reflector and/or the high-frequency regime. Typically, reflectors that are manufactured in panelled sections have an array of strips of tape on the surface. It will be shown numerically that the coupling between the different tapes is negligible, thereby allowing for a superposition of the tape-scattered fields which can be calculated under the assumption that each tape is isolated. This implies that the multiple diffraction between the different strips of tape need not be considered.

An analytical development of the composite tape-diffraction coefficient will be presented. As mentioned previously, the tape-diffraction coefficient is formed by superimposing the fields scattered from both steps which form the boundaries of the tape. The individual step diffraction is constructed from the solutions of two canonical problems; one is the interior modal scattering by a planar short circuit within a simple 2D parallel plate waveguide problem for which the solution is well known, the other is a Wiener–Hopf solution of the plane wave scattering by a semi-infinite ground plane over an infinite ground plane. These solutions will then be combined by the GSMT to form the diffraction coefficient for each of the two steps which form the tape. The two step-scattering solutions will be superimposed as shown in Fig. 3 to form the composite tape-scattering diffraction coefficient.

An $e^{-i\omega t}$ time convention is assumed and suppressed throughout this paper.

2 Tape diffraction coefficient

The fields in the target zone of a compact range consist of the desired reflected geometrical optics (GO) field and the undesired fields which arise from the edges of the reflector as well as from any discontinuity in the reflecting surface, such as an interpanel gap or from a metallic tape covering an interpanel gap (see Fig. 1). The upper and lower reflector edge-diffracted fields in the target zone can be reduced by treating the reflector edges with resistive materials, by serrating the edges, or by using a blended rolled edge. These edge-diffracted fields will not

be addressed here. We will concentrate on the diffracted fields that arise from a surface discontinuity due to perfectly conducting metallic tape on the reflector surface.

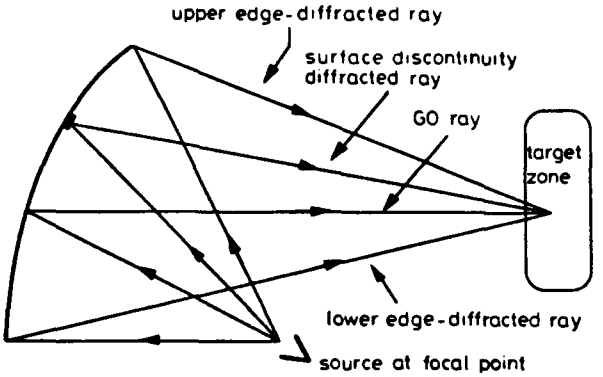


Fig. 1 *Ray mechanisms contributing to the total field in the target zone*

The matallic tape constitutes a perfect electrically conducting (PEC) rectangular cylinder on a parabolic cylinder. However, since the tape width is significantly smaller than the size of the reflector, it will be modelled locally by a PEC rectangular cylinder on an infinite planar surface, as shown in Fig. 2.

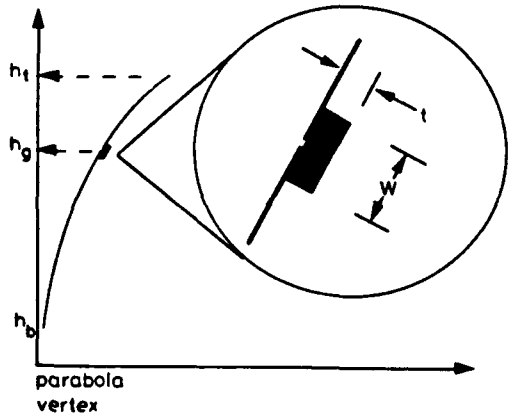


Fig. 2 *Tape geometry*

The solution to the electromagnetic problem in Fig. 3 can be obtained by adding the transverse electric (TE_y) and the transverse magnetic (TM_y) solutions, which are uncoupled. TE_y refers to the case for which the electric field is polarised entirely in the plane of the paper, and similarly TM_y refers to the case for which the magnetic field is polarised entirely in the plane of the paper. Individually, the TE_y and TM_y formulations can be solved as scalar problems.

The steps in the ground plane that form the tape are assumed to be isolated scattering centres. This approximation allows the tape-diffracted fields, U^{scat}_{tape}, to be calculated in two parts, namely by the superposition of the two step-diffracted fields, U^{scat}_{step1} and U^{scat}_{step2}. Thus,

$$U^{scat}_{tape} = U^{scat}_{step1} + U^{scat}_{step2} \quad (1)$$

as shown in Fig. 3, where U represents the scalar y-directed magnetic field for the TE_y polarisation and the scalar y-directed electric field for the TM_y polarisation. In the far field of the tape,

$$U^{scat}_{tape} = U^{i}(Q)\, D^{tape}(\theta^s, \theta^i; t; W)\, \frac{e^{ik\rho}}{\sqrt{\rho}} \quad (2)$$

where $D^{tape}(\theta^s, \theta^i; t; W)$ is the tape diffraction coefficient. The $\{^{TE}_{TM}\}$ designations on the field coefficients are implied and will be omitted for simplicity. It will be shown later that when the isolated scattering centre approximation breaks down it becomes necessary to include multiple

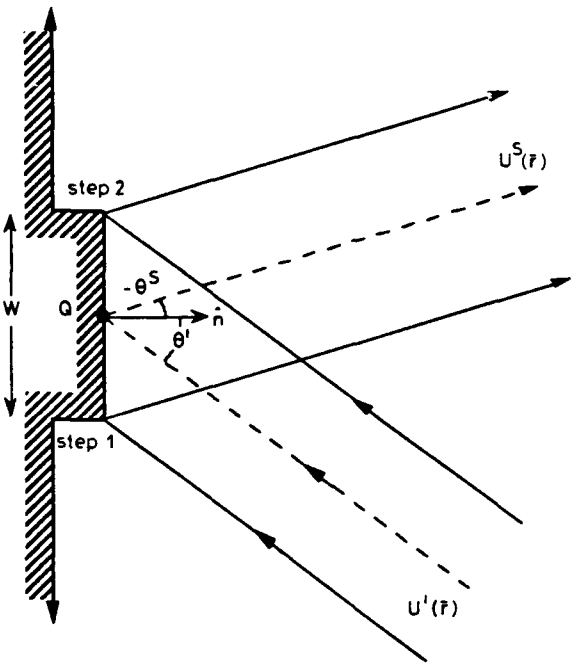


Fig. 3 *Tape-diffracted fields in the far zone formed by superposition of the step-diffracted fields*

step-diffraction terms to account for the step-to-step coupling.

2.1 Step diffraction

The GSMT can be used to resolve the step (Fig. 4) into two less formidable scattering problems provided that a

Fig. 4 *Step in ground plane*

PEC lip of length δ is temporarily introduced to the step, as shown in Fig. 5. As mentioned previously, this problem can be analysed by combining (via the GSMT)

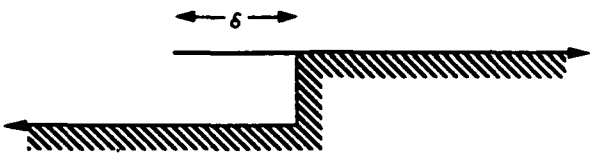


Fig. 5 *Step in ground plane with PEC lip*

the solutions to two canonical problems discussed below and shown in Figs. 6 and 7. A related problem of the diffraction by a thick halfplane in free space was treated by Mittra and Lee [8] and later by Volakis and Ricoy [9].

2.1.1 Canonical problem of scattering by a semi-infinite ground plane over an infinite ground plane: The canonical problem of Fig. 6 can be solved by the Weiner–Hopf technique. Let the exterior and interior regions of Fig. 6 be denoted by regions 1 and 2, respectively. Define the incident, ${}^{\bar{s}(TM/TE)}\bar{\phi}^{inc}_{m}$, and scattered, ${}^{(TM/TE)}\bar{\phi}^{scat}_{n}$, waveguide modes by eqns. 3 and 4, respectively. Note that this structure cannot support an $n =$

0 TM_y mode, but it can support all TE_y modes, including the $n = 0$ mode. So, for the TM_y case, n can take on integer values from 1 to ∞, and, for the TE_y case, n ranges from 0 to ∞. The modal fields are normalised so that all propagating modes contain the same power. The

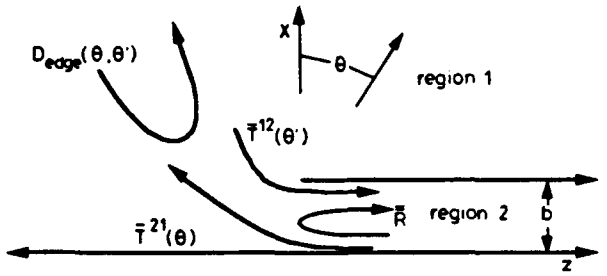


Fig. 6 *PEC semi-infinite ground plane over a PEC infinite ground plane*

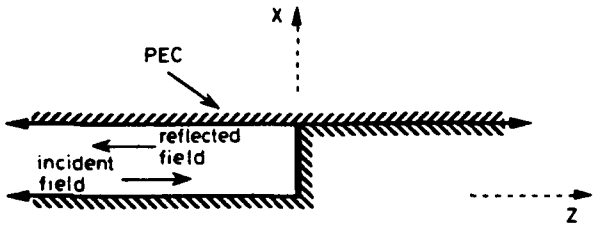


Fig. 7 *Short terminated semi-infinite waveguide below ground plane*

bar above the waveguide modal field symbols signifies the normalisation. There is no physical interpretation of this normalisation for the evanescent modes.

$$ {}^{(TM/TE)}\bar{\phi}_m^{inc} = \frac{1}{\sqrt{(\varepsilon_m \beta_m)}} e^{-i\beta_m z} \begin{Bmatrix} \sin \dfrac{m\pi x}{b} \\ \cos \dfrac{m\pi x}{b} \end{Bmatrix} \tag{3} $$

$$ {}^{(TM/TE)}\bar{\phi}_n^{scat} = \frac{1}{\sqrt{(\varepsilon_n \beta_n)}} e^{i\beta_n z} \begin{Bmatrix} \sin \dfrac{n\pi x}{b} \\ \cos \dfrac{n\pi x}{b} \end{Bmatrix} \tag{4} $$

where

$$ \varepsilon_n = \begin{cases} 1 & \text{if } n \neq 0 \\ 2 & \text{if } n = 0 \end{cases} \tag{5} $$

and

$$ \beta_n = \sqrt{\left[k^2 - \left(\frac{n\pi}{b}\right)^2\right]}; \quad \mathrm{Im}(\beta_n) \geqslant 0 \quad \forall n \tag{6} $$

The reflection of an nth waveguide modal field from the open end of the waveguide when an mth waveguide mode is incident upon the open end is characterised by the modal reflection coefficient R_{mn}. The transmission coefficient which describes the coupling of an incident plane wave field from region 1 into the nth waveguide mode in region 2 is represented by $T_n^{12}(\theta')$, and the transmission coefficient that relates the radiation of the nth modal field into region 1 via the open end is given by $T_n^{21}(\theta)$. Here θ and θ' are measured in a positive sense, i.e. clockwise, from the upward normal to the ground plane. Only one of these transmission coefficients needs to be calculated by the Wiener–Hopf technique since they are related by reciprocity. The coefficient that describes the scattering of the incident plane wave field back into region 1 from the edge discontinuity is given by $D_{edge}(\theta, \theta')$.

The Wiener–Hopf technique is used to calculate the reflection, the transmission and the edge-diffraction coefficients, R_{mn}, $R_n^{21}(\theta)$ and $D_{edge}(\theta, \theta')$, respectively, of the canonical geometry shown in Fig. 6. The particular formulation of the Wiener–Hopf equation is via Jones's method [7, pp. 97–99]. The Fourier transform of ϕ is defined as follows

$$ \hat{\phi}(x, s) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \phi(x, z)\, e^{isz}\, dz \tag{7} $$

where the caret indicates a spectral quantity. Two different types of illumination will need to be considered in this analysis. Specifically, case (*a*) will deal with an interior waveguide modal excitation of the geometry in Fig. 6, and case (*b*) will deal with an external plane wave excitation.

Case (a). Interior waveguide modal excitation: The following field representations are the exact wave number spectral domain fields when the nth waveguide mode (eqn. 3) is incident on the open end

$$ \begin{aligned} {}^{TM}\hat{\phi}_n^{scat}(x, s) = {} & \frac{i}{\sqrt{(2\pi\varepsilon_n \beta_n)}} \frac{n\pi}{b} (-1)^{n+1} \\ & \times \frac{1}{s - \beta_n} L_+^{TM}(\beta_n)\, L_-^{TM}(s) \\ & \times \begin{cases} e^{-\gamma(x-b)} & x > b \\ \dfrac{\sinh(\gamma x)}{\sinh(\gamma b)} & x < b \end{cases} \end{aligned} \tag{8} $$

$$ \begin{aligned} {}^{TE}\hat{\phi}_n^{scat}(x, s) = {} & \frac{i}{\sqrt{(2\pi\varepsilon_n \beta_n)}} (-1)^{n+1} \\ & \times \frac{1}{s - \beta_n} \frac{1}{\gamma} L_+^{TE}(\beta_n)\, L_-^{TE}(s) \\ & \times \begin{cases} -e^{-\gamma(x-b)} & x > b \\ \dfrac{\cosh(\gamma x)}{\sinh(\gamma b)} & x < b \end{cases} \end{aligned} \tag{9} $$

where

$$ \gamma = \sqrt{(s^2 - k^2)} \quad \mathrm{Im}(\gamma) \leqslant 0 \tag{10} $$

and the Weiner–Hopf factors are given by [7, pp. 91–94]

$$ \begin{aligned} L_+^{TM}(s) = {} & i\sqrt{(b)} \sqrt{\left[\frac{\sin(kb)}{kb}\right]} \\ & \times \exp\left[\frac{ibs}{\pi}\left(1 - c + \ln\left(\frac{2\pi}{kb}\right) + i\frac{\pi}{2}\right)\right] \\ & \times \exp\left[\frac{ib\gamma}{\pi} \ln\left(\frac{s - \gamma}{k}\right)\right] \\ & \times \prod_{l=1}^{\infty} \left(1 + \frac{s}{\beta_l}\right) \exp\left(i\frac{sb}{l\pi}\right) \\ = {} & L_-^{TM}(-s) \end{aligned} \tag{11} $$

and

$$ L_+^{TE}(s) = i\,(s + k)\, L_+^{TM}(s) = L_-^{TE}(-s) \tag{12} $$

where

$$ c = 0.057721\ldots \quad \text{(Euler's constant)} \tag{13} $$

These spectral solutions of the radiated (transmitted) and reflected fields, eqns. 8 and 9, must be converted to the spatial domain by the inverse Fourier transform. This results in a contour integral representation of the configuration space field.

The complex s-plane topologies of the radiated and reflected fields are shown in Figs. 8 and 9, respectively. With the exception of the transverse electromagnetic (TEM) waveguide mode pole, the topologies of the two polarisations are similar.

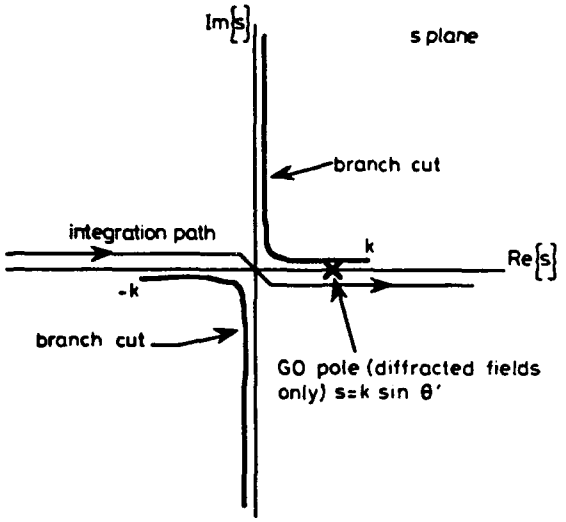


Fig. 8 *s-plane representation of the fields radiated into region 1 due to an incident waveguide mode and the fields diffracted back into region 1 due to an incident plane wave arriving at an angle θ'*

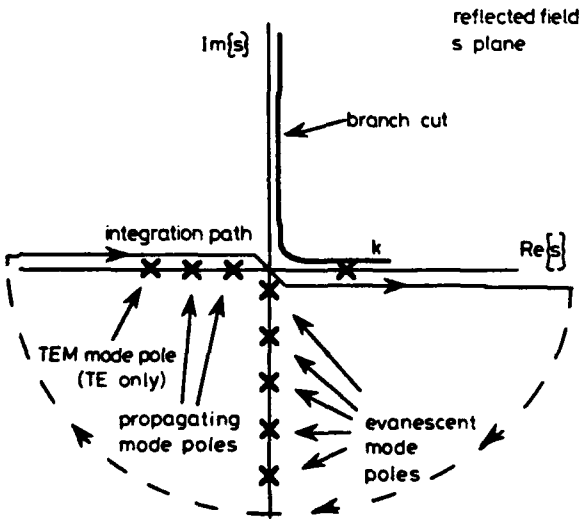


Fig. 9 *s-plane representation of fields reflected back into region 2 due to an incident waveguide mode*

The contour integration for the scattered fields for $x < b$ due to a waveguide mode excitation is shown in Fig. 9. For $z > 0$, which corresponds to fields reflected back into the waveguide interior, the contour is closed in the lower halfplane at infinity (dashed contour) to facilitate the use of the residue theorem. It can be shown that this dashed contour does not contribute to the integral, thereby allowing the reflected fields in the waveguide to be equated to the residue contributions which are precisely the waveguide modal fields. By inspection of the reflected fields in the spatial domain, the reflection coefficient can be extracted and is given in eqns. 14 and 15.

$$R_{mn}^{TM} = mn\left(\frac{\pi}{b}\right)^2 (-1)^{m+n+1} \frac{1}{\beta_m + \beta_n} \times \frac{1}{b\sqrt{(\beta_m \beta_n \varepsilon_m \varepsilon_n)}} L_+^{TM}(\beta_m)\, L_+^{TM}(\beta_n) \tag{14}$$

$$R_{mn}^{TE} = (-1)^{m+n+1} \frac{1}{\beta_m + \beta_n} \frac{1}{b\sqrt{(\beta_m \beta_n)}} \times L_+^{TE}(\beta_m)\, L_+^{TE}(\beta_n) \tag{15}$$

The radiated wave number spectra of the fields in region 1 contain a branch cut, as does the field in any infinite region, which prevents an exact evaluation of the inverse transform (see Fig. 8). However, the far field can be well approximated by evaluating the integral using the method of steepest descent. Note that, in the spectra of the radiated field, eqns. 8 and 9, the apparent pole at $s = \beta_n$ due to the $1/(s - \beta_n)$ term is removed by the zeros of the $L_-^{TM}(s)$ or $L_-^{TE}(s)$ functions.

Before the steepest descent approximation is applied, it is convenient to make the following transformation to the angular spectrum

$$s = k \sin \omega \tag{16}$$

along with the following substitutions

$$x - b = \rho \cos \theta \tag{17}$$

and

$$z = \rho \sin \theta \tag{18}$$

Under the transformation given in eqn. 16, the s-plane is transformed to the ω-plane. This ω-plane is shown in Fig. 10. The ω-plane representation is convenient to perform the steepest descent approximation since all the branch cuts have been removed. The steepest descent path (SDP), the saddle point, and the contour C onto which the original s-plane integration path was mapped are indicated in Fig. 10. After performing the steepest descent analysis and inspecting the solution, the transmission coefficients can be extracted. Note that the solution is very efficient since it contains only elementary functions and is in closed form.

$$T_n^{21\,TM}(\theta) = \frac{i}{\sqrt{(2\pi)}} \frac{n\pi}{b} (-1)^n \frac{\sqrt{(k)\cos\theta}}{k \sin\theta + \beta_n} L_+^{TM}(\beta_n) \times L_+^{TM}(k \sin\theta) \frac{1}{\sqrt{(\beta_n)}} \exp\left(-i\frac{\pi}{4}\right) \tag{19}$$

$$T_n^{21\,TE}(\theta) = \frac{1}{\sqrt{(2\pi k)}} (-1)^n \frac{1}{k \sin\theta + \beta_n} L_+^{TE}(\beta_n) \times L_+^{TE}(k \sin\theta) \frac{1}{\sqrt{(\beta_n \varepsilon_n)}} \exp\left(-i\frac{\pi}{4}\right) \tag{20}$$

Case (b). External plane wave excitation: To find $D_{edge}(\theta, \theta')$, we need to determine the edge scattering by

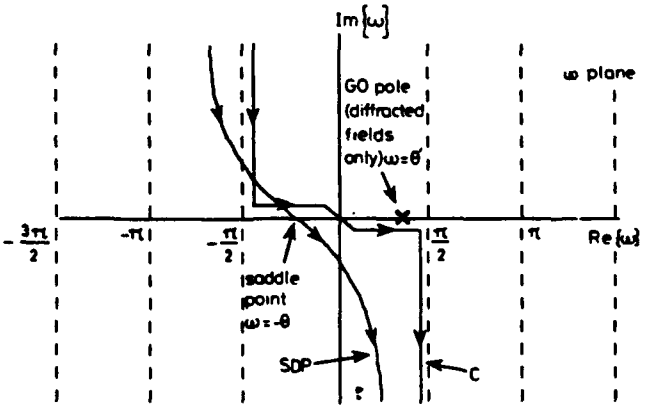


Fig. 10 *ω-plane representation of the fields radiated into region 1 due to an incident waveguide mode and the fields diffracted back into region 1 due to incident plane wave arriving at an angle θ'*

an incident plane wave field as shown below

$$^{(TM/TE)}\phi^{inc}(\theta') = e^{-ik((x-b)\cos\theta' + z\sin\theta')} \quad (21)$$

with an assumed form of the reflected geometrical optics (GO) field given by

$$^{(TM/TE)}\phi^{ref}(\theta') = \mp e^{-ik(-(x-b)\cos\theta' + z\sin\theta')} \quad (22)$$

The scattered field is the difference between the total field and the incident field plus the assumed reflected field. The wave number spectrum of the edge-scattered field for $x > b$ is given by

$$^{TM}\hat{\phi}^{scat}_{edge}(x > b, s; \theta') = -\frac{2k}{\sqrt{(2\pi)}}\frac{\cos\theta'}{s - k\sin\theta'} L_+^{TM}(k\sin\theta') \times L_-^{TM}(s)\, e^{-\gamma(x-b)} \quad (23)$$

and

$$^{TE}\hat{\phi}^{scat}_{edge}(x > b, s; \theta') = -i\sqrt{\left(\frac{2}{\pi}\right)}\frac{1}{s - k\sin\theta'}\frac{1}{\gamma} L_+^{TE}(k\sin\theta') \times L_-^{TE}(s)\, e^{-\gamma(x-b)} \quad (24)$$

The scattered field spectrum must be converted to the spatial domain by the inverse Fourier transform The complex s-plane topology for both polarisations of the scattered (diffracted) field is also shown in Fig. 8. The residue of the pole at $s = k\sin\theta'$ shown in eqns. 23 and 24 is the GO contribution which corrects the assumed reflected field given in eqn. 22. If the GO pole (at $\omega = \theta'$) in the diffracted field ω-plane (Fig. 10) is to the left of the saddle point, then it is necessary to include the residue contribution of the pole because it would be enclosed by the SDP and C. The diffracted fields in region 1 can be evaluated asymptotically by the method of steepest descent in a similar manner to the radiated fields in case (*a*). For convenience, the spectra will also be transformed into the angular domain by eqn. 16, along with the substitutions given in eqns. 17 and 18. The corresponding diffracted field ω-plane is shown in Fig. 10.

By inspection of the steepest descent solution, the edge-diffraction coefficients can be expressed as

$$D^{TM}_{edge}(\theta, \theta') = \sqrt{\left(\frac{2k}{\pi}\right)}\frac{\cos\theta\cos\theta'}{\sin\theta + \sin\theta'} L_+^{TM}(k\sin\theta) \times L_+^{TM}(k\sin\theta')\exp\left(-i\frac{\pi}{4}\right) \quad (25)$$

$$D^{TE}_{edge}(\theta, \theta') = -\sqrt{\left(\frac{2}{k^3\pi}\right)}\frac{1}{\sin\theta + \sin\theta'} L_+^{TE}(k\sin\theta) \times L_+^{TE}(k\sin\theta')\exp\left(-i\frac{\pi}{4}\right) \quad (26)$$

As mentioned earlier, $T_n^{12(TM/TE)}(\theta)$ is related to $T_n^{21(TM/TE)}(\theta)$ by reciprocity, as will be shown later in eqn. 32.

2.1.2 Canonical problem of reflection by a planar short within a semi-infinite parallel plate waveguide: The second canonical problem is shown in Fig. 7. The solution for this case can be found in many elementary electromagnetic textbooks, and therefore it will be simply stated here without proof. The excitation is assumed to be a waveguide mode travelling to the right, $^{(TM/TE)}\phi_n^{inc}$, which is incident on a planar short at $z = 0$ (see Fig. 7). The only reflected modal field is given by $^{(TM/TE)}\phi_n^{refl}$. The ratio of $^{(TM/TE)}\phi_n^{refl}(z = 0)/^{(TM/TE)}\phi_m^{inc}(z = 0)$, the reflection coefficient, is given by $^{(TM/TE)}\Gamma = \mp\delta_{mn}$, where δ_{mn} is the Kronecker delta function

$$\delta_{mn} = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases} \quad (27)$$

2.1.3 Generalised scattering matrix formulation: The scattering matrices required in the GSMT procedure can be defined in terms of the reflection, transmission and diffraction coefficients pertaining to the two canonical problems as follows [7, pp. 207–210]

$$S_{11}^{(TM/TE)} = D_{edge}^{(TM/TE)}(\theta, \theta') \quad (28)$$

$$[S_{22}^{(TM/TE)}] = [R_{mn}^{(TM/TE)}] \quad (29)$$

$$[S_\Gamma^{(TM/TE)}]_{mn} = \mp\delta_{mn} \quad (30)$$

$$[S_{21}^{(TM/TE)}] = [T_n^{21(TM/TE)}(\theta)] \quad (31)$$

and by reciprocity it can be shown that

$$[S_{12}^{(TM/TE)}] = \pm\frac{\sqrt{(8\pi ik)}}{b}[S_{21}^{(TM/TE)}]^T \quad (32)$$

When implementing the solution on the computer, it is not possible to include the effects of an infinite number of evanescent modes since that would require the matrices to be of infinite order. For solutions of engineering accuracy, it is typically sufficient to use scattering matrices of order $N \times N$, where N is the number of propagating modes plus 5.

The total scattering coefficient, $S_{scat}^{region1}$, accounting for all of the multiple interactions, is given below. The $\{TM/TE\}$ superscripts are omitted for notational simplicity; however, they are implied

$$S_{scat}^{region1}(\delta) = S_{11} + [S_{21}][P][S_\Gamma][P]\cdot \cdot([I] - [S_{22}][P][S_\Gamma][P])^{-1}[S_{12}] \quad (33)$$

In eqn. 33, $[P]$, the transmission line scattering matrix, is independent of polarisation

$$[P]_{mn} = e^{i\beta_n\delta}\delta_{mn} \quad (34)$$

β_n is the propagation constant of the nth mode of the parallel plate waveguide, δ is the length of the PEC lip.

The step-diffraction coefficient is recovered by taking the limit as $\delta \to 0$ of the scattered field solution in region 1 (eqn. 33)

$$^{(TM/TE)}D^{step}(\theta^s, \theta^i; t) = S_{11} + [S_{21}][S_\Gamma]([I] - [S_{22}][S_\Gamma])^{-1}[S_{12}] \quad (35)$$

The first term in this expression is due to the scattering from the edge of the semi-infinite ground plane over an infinite ground plane, and the second term is due to the energy that is coupled into the waveguide of vanishing length and is then reradiated back into region 1. The total step-diffracted field, u^d, can be calculated if the incident field, u^i, at the discontinuity is known. Thus,

$$u^d = u^i\, ^{(TM/TE)}D^{step}(\theta^s, \theta^i; t)\frac{e^{ik\rho}}{\sqrt{\rho}} \quad (36)$$

where ρ is the distance from the edge to the observation point. If the polarisation of interest is TM_y, then u^d and u^i represent the y-directed electric field. Similarly, for TE_y, u^d and u^i represent the y-directed magnetic field.

2.2 First-order diffraction coefficients for a tape on the compact-range reflector

To facilitate the locally planar approximation of the tape and the ground plane we assume a plane wave field incident on the tape. Since the width of the tape is very small

compared to the distance from the tape to any point in the target zone, we can also assume that the rays diffracted from the two steps that form the tape arriving at any point in the target zone are parallel. Therefore, the composite tape-diffraction coefficient can be calculated as follows

$$D^{tape}(\theta^s, \theta^i; t; W) = \exp\left[-ik\frac{W}{2}(\sin\theta^i + \sin\theta^s)\right] D^{step}(\theta^s, \theta^i; t) + \exp\left[ik\frac{W}{2}(\sin\theta^i + \sin\theta^s)\right] \times D^{step}(\pi - \theta^s, \pi - \theta^i; t) \quad (37)$$

where θ^s and θ^i are the scattered and incident field directions, respectively, measured from the normal as shown in Fig. 3. Since the target zone is in the far zone of the tape-scattered fields, we can ignore the GO tape-scattered field (the spectral width of the GO tape-scattered field shrinks to zero in the far field).

2.3 *Obtaining the 3D tape-scattered fields in the target zone by the equivalent current method*

The tape-scattered fields in the target zone of the reflector can be obtained via the equivalent current method (ECM) [2–4] or the incremental length diffraction coefficients [5, 6] by using the 2D tape-diffraction coefficient developed in Section 2.2. The idea behind the ECM is to determine the locally tangent equivalent free space electric (I) and magnetic (M) currents along the midline (centre) of the upper face of the 3D tape. The currents are determined by assuming that the tape is two dimensional and that the direction of incidence and diffraction make an angle, β_0' and β_0, respectively with the vector tangent to the length of the tape. It is noted that β_0' and β_0 are functions of position along the length of the tape [3, 4]. Thus,

$$\begin{Bmatrix} I(Q_E) \\ M(Q_E) \end{Bmatrix} = \frac{-\exp\left(i\frac{\pi}{4}\right)}{\sqrt{[\sin(\beta_0')\sin(\beta_0)]}} \sqrt{\left(\frac{8\pi}{k}\right)} \begin{Bmatrix} Y_0 \\ Z_0 \end{Bmatrix} \hat{e} \cdot \begin{Bmatrix} E^{inc}(Q_E)\,{}^{TM}D^{tape} \\ H^{inc}(Q_E)\,{}^{TE}D^{tape} \end{Bmatrix} \frac{1}{\sqrt{[\sin(\beta_0')\sin(\beta_0)]}} \quad (38)$$

where Q_E is an illuminated point on the tape, $\hat{e}$ is the local unit tangent to the length of the tape, and Y_0 and Z_0 are the free space admittance and impedance, respectively.

These equivalent currents are then inserted into a radiation integral which integrates along the entire directly illuminated finite length of the tape

$$E_{3D}^{tape}(P) \sim \frac{ikZ_0}{4\pi} \int_{tape} [\hat{R} \times \hat{R} \times I\hat{e} + Y_0 \hat{R} \times M\hat{e}] \frac{e^{ikR}}{R} dl' \quad (39)$$

P is the observation point in the target zone, R is the vector from Q_E to P, and here $\hat{R}$ is the unit vector in the R direction. Since the radiation integral is finite in the spatial domain, diffraction from the endpoints of the tape are inherent in this formulation under the assumption that the endpoints of the tape do not perturb the equivalent currents.

3 Numerical results

The cases presented in this section simulate the tape-scattered fields from a 2D compact range with a focus 7.62 m (25 ft) and a target zone 12.19 m (40 ft) from the vertex of the parabola. All data are normalised by the GO fields of a perfect infinite parabola. The solid curves result from an application of the theory presented in this paper. The reference solution shown by the dotted lines was formed by subtracting a method of moments (MoM) solution of the scattered fields for a finite smooth (no tape) parabola from the MoM computed scattered fields for the same parabola with a metallic tape on the surface. This subtraction process attempts to 'cancel' all but the tape-scattered fields, and does so very well. However, at low levels in the TE$_z$ polarisation case there is some noticeable interaction between the tape and the end of the parabola that does not cancel completely in the subtraction. This interaction manifests itself in an oscillation about the true tape-scattered fields.

Fig. 11 shows the TM$_z$ tape-scattered fields in the target zone for three different tape widths and three different tape thicknesses. In all cases, the agreement is very good and seems to improve as the tape becomes wider. This is consistent with the fact that the steps interact with each other less and less as the tape widens. Note the trend of an increase of 6 dB/(doubling of tape width).

Fig. 12 examines a tape width of 10.16 cm (4 in) with a thickness of 0.254 cm (0.1 in) at three different frequencies. The agreement is again excellent, and it appears to improve as the frequency increases. This corresponds to the tape becoming electrically wider and hence multiple diffraction (which is not included) becoming less significant.

When multiple tapes are placed on the same reflector, they will interact with each other (on the tape level as opposed to the step-to-step level). Fig. 13 attempts to put these multiple tape interactions into perspective. The MoM solution was performed with both tapes present and therefore contains all higher-order tape interactions. The present solution is a superposition of the first-order tape-diffracted fields. Since the first-order analysis for this size tape ($w = 10.16$ cm (4 in), $t = 0.254$ cm (0.1 in) at $f = 1$ GHz) was shown in the previous figures to be quite accurate, Fig. 13 is a measure of the tape-to-tape interaction. The tape-scattered fields are plotted for various vertical separations of 0.305 m (1 ft), 0.914 m (3 ft), and 1.83 m (6 ft). For all three cases the two solutions track remarkably well implying that, for TM$_z$ polarisation, tape-to-tape interactions are not significant.

The first-order results for the TE$_z$ polarisation is not as promising as the first-order results for the TM$_z$ polarisation (see Fig. 14). In this example we are presenting comparisons for three different widths: 1.35λ, 0.677λ, and 0.339λ. It is obvious that, compared to the TM$_z$ case, the step-to-step interactions are high, resulting in poor agreement with the reference solution. This is because the step-scattered grazing fields for the TM$_z$ polarisation must vanish, yielding small step-to-step coupling, whereas the TE$_z$ polarisation fields do not vanish on the boundary so the step-to-step coupling is more significant. The TE$_z$ polarisation would benefit greatly by including the multiply step-diffracted fields which will be considered in the future.

4 Conclusion

We have presented an analytic solution to first order to predict the scattering of a metallic tape on a compact-range reflector when it is observed in the target zone. Since it remains accurate at the tape reflection shadow boundary, it is a uniform solution.

It was shown by comparison with a reference MoM solution that the TM_z polarisation solution yields excellent results, whereas the TE_z polarisation would benefit from the inclusion of multiple step-diffraction mechanisms. It was also shown empirically that the tape-to-tape interactions for the TM_z polarisation are not significant and can be ignored.

5 Acknowledgments

The authors are indebted to Dr. I.J. Gupta for numerous discussions concerning the application of the tape-diffraction coefficient to the compact range. We would also like to express gratitude to Ms. J. Feil for providing the MoM reference solution and to Dr. R.J. Burkholder

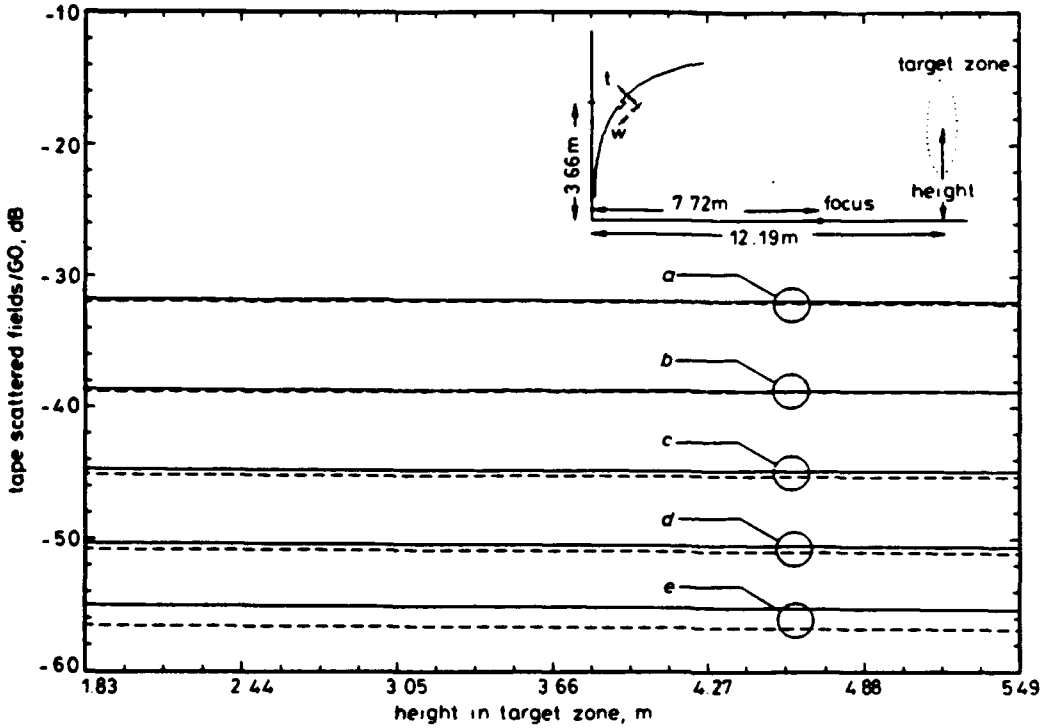


Fig. 11 *Normalised TM_z tape-scattered fields for various tape widths and thicknesses*

$f = 1$ GHz
——— present solution
------- moment method
(a) $t = 3.39 \times 10^{-3}\lambda$, $w = 0.339\lambda$
(b) $t = 1.69 \times 10^{-3}\lambda$, $w = 0.339\lambda$
(c) $t = 8.47 \times 10^{-3}\lambda$, $w = 0.339\lambda$
(d) $t = 8.47 \times 10^{-3}\lambda$, $w = 0.169\lambda$
(e) $t = 8.47 \times 10^{-3}\lambda$, $w = 8.47 \times 10^{-3}\lambda$

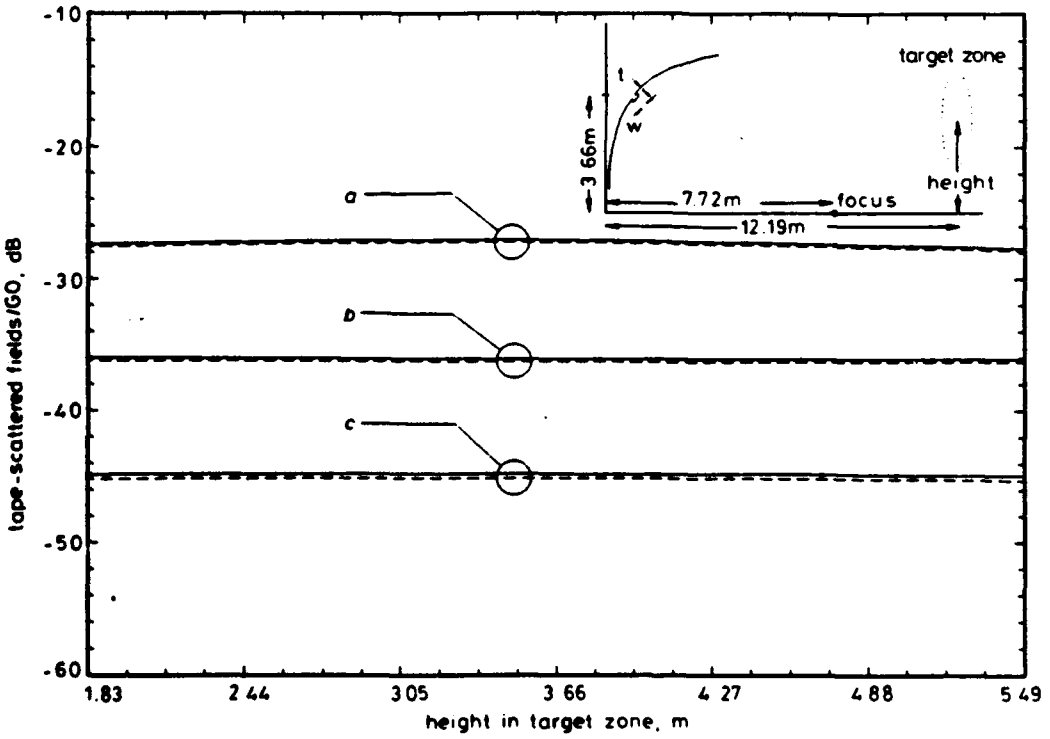


Fig. 12 *Normalised TM_z tape-scattered fields at various frequencies*

——— present solution
------- moment method
$w = 10.16$ cm, $t = 0.254$ cm
(a) $f = 4$ GHz
(b) $f = 2$ GHz
(c) $f = 1$ GHz

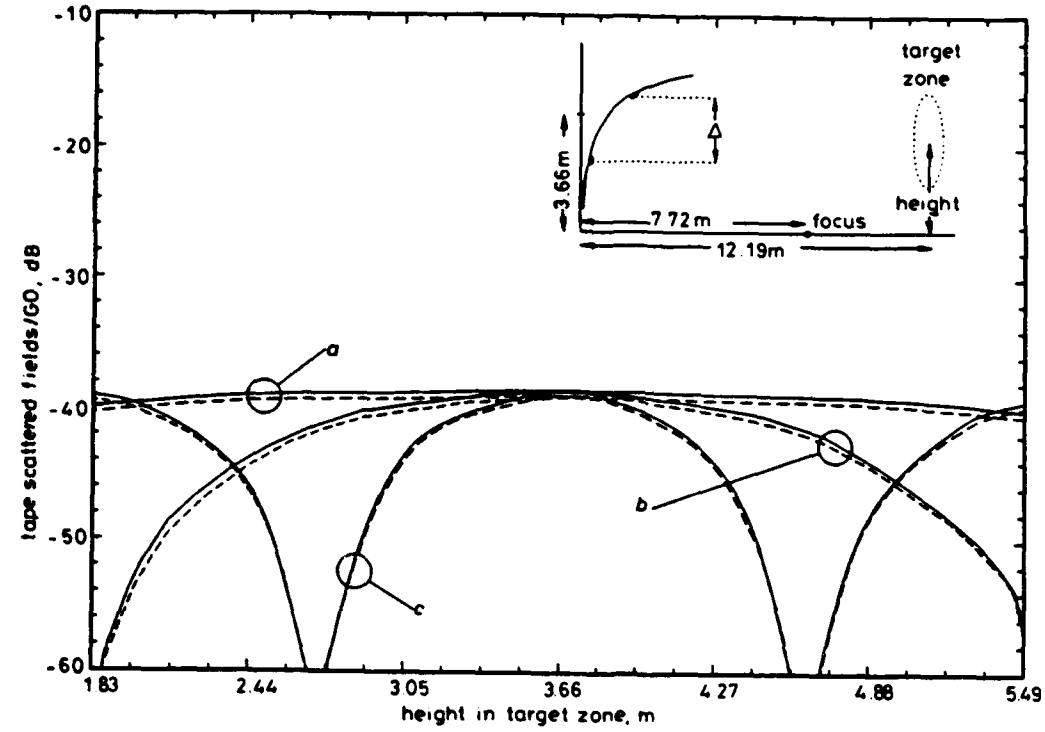


Fig. 13 *Normalised TM, multiple tape-scattered fields for various tape separations*

——— present solution
------- moment method
$w = 0.339\lambda$, $t = 8.47 \times 10^{-3}\lambda$, $f = 1$ GHz
(*a*) $\Delta = 0.3$ m
(*b*) $\Delta = 0.9$ m
(*c*) $\Delta = 1.8$ m

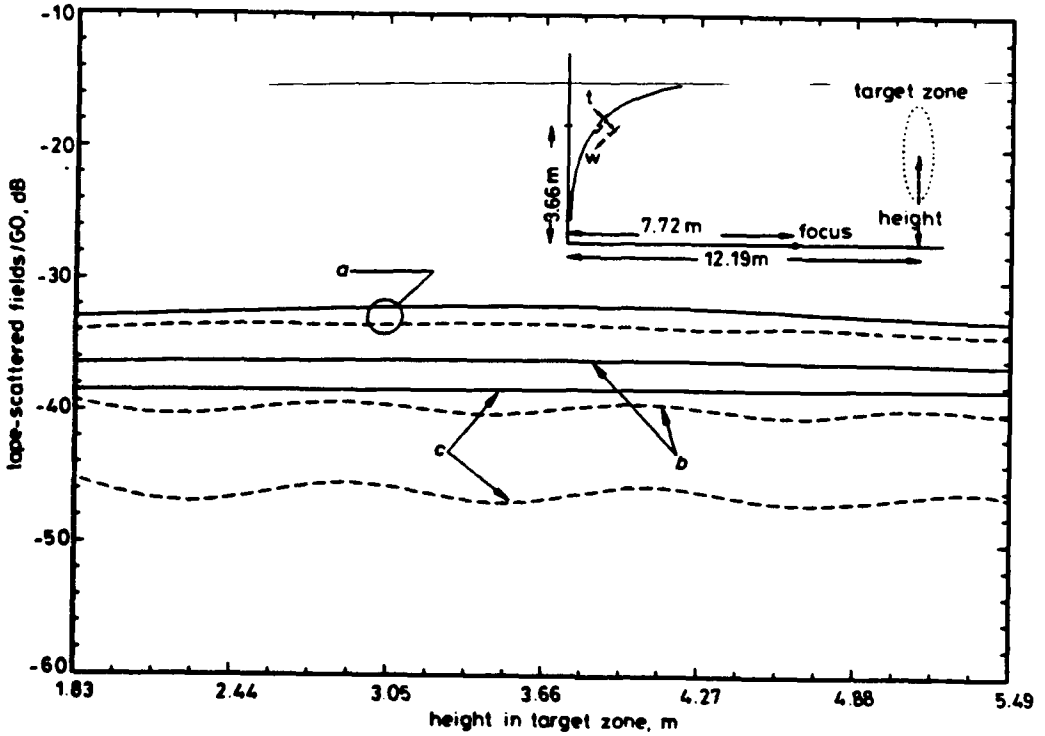


Fig. 14 *Normalised TE, tape-scattered fields for various tape widths*

——— present solution
------- moment method
$t = 8.47 \times 10^{-3}\lambda$, $f = 1$ GHz
(*a*) $w = 1.35\lambda$
(*b*) $w = 0.677\lambda$
(*c*) $w = 0.339\lambda$

for supplying the initial version of the GSMT code. This work was supported in part by an Ohio Aerospace Institute/NASA Space Grant Doctoral Fellowship and in part by the Joint Services Electronics Program under contract N00014-89-J-1007.

6 References

1 GUPTA, I.J., and BURNSIDE, W.D.: 'Scattered fields of metallic tapes used to cover gaps in compact range reflectors'. AMTA Conference, pp. 15-35–15-39, Monterey, CA, Oct. 1989
2 RYAN, C.E. Jr., and PETERS, L. Jr.: 'Evaluation of edge diffracted fields including equivalent currents for caustic regions', *IEEE Trans.*, 1970, AP-17, pp. 292–299
3 PATHAK, P.H.: 'Techniques for high frequency problems', *in* LO, Y.T., and LEE, S.W. (Eds.): 'Antenna handbook, theory, application and design' (Van Nostrand Reinhold, 1988), pp. 4-96–4-102
4 KNOTT, E.F., and SENIOR, T.B.A.: 'Comparison of three high-frequency diffraction techniques', *Proc. IEEE*, 1974, 62, pp. 1468–1478
5 MITZNER, K.M.: 'Incremental length diffraction coefficients'. Aircraft Division, Northrop Corp. Tech. Report AFAL-TR-73-296, 1974
6 SHORE, R.A., and YAGHJIAN, A.D.: 'Incremental diffraction coefficients for planar surfaces', *IEEE Trans.*, 1988, AP-36, (1), pp. 55–70
7 MITTRA, R., and LEE, S.W.: 'Analytical techniques in the theory of guided waves' (The Macmillan Company, New York, 1971)
8 LEE, S.W., and MITTRA, R.: 'Diffraction by thick conducting half-plane and a dielectric-loaded waveguide', *IEEE Trans.*, 1968, AP-16, (4), pp. 454–461
9 VOLAKIS, J.L., and RICOY, M.A.: 'Diffraction by a thick perfectly conducting half-plane', *ibid.*, 1987, AP-35, (1), pp. 62–72

Journal of Electromagnetic Waves and Applications, Vol. 6, No. 5/6, 733-750, 1992

Integral Equations for the Scattering by a Three Dimensional Inhomogeneous Chiral Body

R.G. Rojas

The ElectroScience Laboratory
Department of Electrical Engineering
The Ohio State University
Columbus, OH 43212
USA

Abstract– Integral equations are obtained for the electromagnetic (EM) scattering by an inhomogeneous, isotropic, three dimensional chiral body. The chiral body is assumed to be in free space, and it can be attached to a perfect electric conducting (PEC) body. The integral equations are obtained with the help of vector-dyadic identities and the free space dyadic Green's function. These equations are expressed in terms of a volume integral with the electric field as the unknown and surface integrals where the tangential components of the electric field and its curl are the unknowns. The integral equations are then transformed into a linear system of simultaneous equations by means of the moment method technique. Expressions for the scattered field in the far-zone are also obtained by replacing the dyadic Green's function and its curl with their large argument approximations. Furthermore, closed form expressions are obtained for the fields and dipole moments induced inside an electrically small, homogeneous chiral sphere where it is assumed that the fields are constant. Finally, closed form expressions and numerical results for the fields scattered by the small chiral sphere and its bistatic echo area are also obtained.

I. INTRODUCTION

The interaction of electromagnetic fields at microwave frequencies with the optically active chiral media has attracted the attention of the electromagnetic community due to its potential applications in the field of antennas, microwave devices, waveguide propagation, scattering, etc [1–5]. For example, in scattering applications it may be possible to use chiral material to coat a scatterer and thus control its scattering properties more efficiently than a regular dielectric coating due to the extra degree of freedom offered by the presence of the chiral parameter. The purpose of this paper is to develop a set of integral equations for the scattering of EM fields by a three dimensional inhomogeneous chiral body in free space which can be attached to another perfect electric conducting (PEC) body as shown in Fig. 1. Note that a well known procedure known as the volume equivalent current method [6], where the inhomogeneous material scatterer (in this case a chiral body) is replaced by a set of unknown equivalent currents, can be used to solve for the fields scattered by the chiral body due to an incident field. The integral equations for the unknown equivalent electric $\overline{J}_{eq}$ and magnetic $\overline{M}_{eq}$ currents are easily obtained with this method; however, these equations are coupled. For

example, for a chiral body in free space without the presence of any other scatterers, the volume equivalent current method yields a set of equations where there are six scalar equations with six unknowns. Furthermore, all the integrals are volume integrals for a three dimensional body, and surface integrals for a two-dimensional body. This method has been used by Kluskens and Newman [7] to solve a two dimensional problem. Note that for the simpler case of a homogeneous chiral body, a formulation which only involves surface currents, can be used as was done by Morita [8] for a dielectric scatterer. The goal of this paper is then to obtain a system of integral equations for three dimensional bodies in terms of a volume integral with as few unknowns as possible and a set of surface integrals also with as few unknowns as possible [9]. The starting point of the present paper are Maxwell's equations and the constitutive relations for chiral media. For the reader who is interested in a detailed explanation of these equations and the various theoretical aspects of EM field propagation inside chiral media, there are many papers in the literature which deal with these subjects [10–23]. It is noted that this list of papers is by no means complete, but it is representative of the work that is presently being conducted in this area. Different expressions (all of them equivalent) exist for the constitutive relations; however, a particular set of equations, popularized by Bohren [10–12], is used here. These constitutive relations are very convenient for the derivation of the integral equations because they satisfy simple duality relations. The paper is organized as follows. In Section II, the integral equations for a three dimensional chiral body attached to a PEC body are developed. These equations are obtained by transforming the vector differential wave equation for the electric field into a set of integral equations with the help of various vector-dyadic identities and the free space dyadic Green's function. It is noted that these equations can also be obtained by starting with the equivalent polarization current formulation. The procedure followed here to obtain the integral equations is similar to that used by Tai [24] for a dielectric body. In Section III, the integral equations developed in Section II are transformed into a linear system of simultaneous equations by means of the moment method procedure. In Section IV, an expression for the scattered field in the far-zone is obtained by only keeping terms of $O(R_0)$ where R_0 is the distance from the origin to the observation point. In Section V, closed form expressions using the integral equations developed in Section II are obtained for the fields and dipole moments induced inside a small homogeneous chiral sphere. This calculation can be accomplished because the sphere is assumed to be much smaller than the two wavelengths inside the material (right and left circularly polarized) such that the induced fields are constant. Once the electric and magnetic dipole moments are computed, expressions for the far-zone scattered fields and the bistatic echo area (or radar cross-section) are developed. Simple numerical results are also computed for the chiral sphere. Note that the purpose of considering the scattering problem in Section V is to verify the validity of the integral equations developed in Section II. Finally, in Section VI some concluding remarks are given. Note that in the analysis that follows, a $\exp(-i\omega t)$ time dependence for the fields is assumed and suppressed.

II. INTEGRAL EQUATIONS FOR THREE DIMENSIONAL BODIES

For time harmonic fields, Maxwell's equations inside the source-free chiral body of Fig. 1 can be written as

$$\nabla \times \overline{E} = i\omega \overline{B} \quad ; \quad \nabla \times \overline{H} = -i\omega \overline{D} \tag{1}$$

whereas the constitutive relations are given by [1]

$$\overline{D} = \epsilon \overline{E} + i\gamma \overline{B} \quad ; \quad \overline{H} = i\gamma \overline{E} + \frac{1}{\mu} \overline{B} \tag{2}$$

where $\epsilon(\overline{r})$, $\mu(\overline{r})$, and $\gamma(\overline{r})$ are the permittivity, permeability, and chiral admittance, respectively. Note that all three parameters are functions of position for an inhomogeneous chiral medium and are in general complex. An alternative set of equations to those given in (1) and (2) exist which are more convenient for the present analysis. That is, (1) and (2) can be rewritten as

$$\nabla \times \overline{E} = \frac{i\omega\mu_c}{\tau} \overline{H} + \frac{k^2}{\tau} \beta \overline{E} \tag{3a}$$

$$\nabla \times \overline{H} = -\frac{i\omega\epsilon}{\tau} \overline{E} + \frac{k^2\beta}{\tau} \overline{H} \tag{3b}$$

where $k = \omega\sqrt{\epsilon\mu_c}$ and [19]

$$\gamma = \omega\epsilon\beta \quad ; \quad \mu = \frac{\mu_c}{\tau} \quad ; \quad \tau = 1 - k^2\beta^2 \tag{3c}$$

It should be emphasized that the parameter k above is not the wave number of the chiral medium. The fields $\overline{D}$ and $\overline{B}$ can then be expressed as

$$\overline{D} = \epsilon \left(\overline{E} + \beta \nabla \times \overline{E}\right) \quad ; \quad \overline{B} = \mu_c \left(\overline{H} + \beta \nabla \times \overline{H}\right) \tag{4}$$

Note that (3*a*), (3*b*), and (4) satisfy the following duality transformations:

$$\begin{aligned} &\overline{E} \to \overline{H} \quad , \quad \overline{H} \to -\overline{E} \quad , \quad \epsilon \to \mu_c \quad , \quad \mu_c \to \epsilon \quad , \quad \beta \to \beta \\ &\overline{D} \to \overline{B} \quad \text{and} \quad \overline{B} \to -\overline{D} \end{aligned} \tag{5}$$

That is, (3*b*) can be obtained from (3*a*) and the expression for $\overline{B}$ in (4) can also be obtained from the expression for $\overline{D}$ by applying (5).

A vector differential equation can be obtained for the electric field $\overline{E}$ by appropriately combining (3*a*) and (3*b*), namely

$$\nabla \times \left(\frac{\tau}{\mu_c} \nabla \times \overline{E}\right) - \omega^2 \nabla \times \left(\epsilon\beta \overline{E}\right) - \omega^2 \epsilon\beta \nabla \times \overline{E} - \omega^2 \epsilon \overline{E} = \overline{0} \tag{6a}$$

By the duality transformation of (5), the following equation for $\overline{H}$ can be obtained from (6*a*)

$$\nabla \times \left(\frac{\tau}{\epsilon} \nabla \times \overline{H}\right) - \omega^2 \nabla \times \left(\mu_c\beta \overline{H}\right) - \omega^2 \mu_c\beta \nabla \times \overline{H} - \omega^2 \mu_c \overline{H} = \overline{0} \tag{6b}$$

Equations (6*a*) and (6*b*) reduce to the well known vector wave equations when the chiral medium becomes homogeneous. Before the integral equations are obtained, it is necessary to specify the boundary conditions satisfied by the fields at the boundaries S_b, S_c, and S_d depicted in Fig. 1. Taking into account that the tangential components of the electric $\overline{E}$ and magnetic $\overline{H}$ fields are continuous

on S_b and that the tangential components of $\overline{E}$ vanish on a PEC surface, the fields $\overline{E}_a$ and $\overline{E}_b$ satisfy the following boundary conditions (where $\overline{E}_a$ and $\overline{E}_b$ are the electric fields inside and outside the chiral body, respectively)

$$\hat{n}\times\overline{E}_a=\hat{n}\times\overline{E}_b \quad ; \quad \hat{n}\times\nabla\times\overline{E}_b=\frac{\mu_0}{\mu_c}\tau\hat{n}\times\nabla\times\overline{E}_a-\omega^2\mu_0\epsilon\beta\hat{n}\times\overline{E}_a \text{ on } S_b \quad (7a)$$

$$\hat{n}\times\overline{E}_b=0 \quad \text{on} \quad S_c \text{ and } \quad \hat{n}\times\overline{E}_a=0 \quad \text{on} \quad S_d \quad (7b)$$

where $\hat{n}$ is the unit vector normal to the surfaces $S_\rho(\rho=b,c,d)$. Note that S_b is the boundary between the chiral body and free space, S_c is the surface of the PEC body not in contact with the chiral body, and S_d is the surface of the PEC body attached to the chiral body (see Fig. 1).

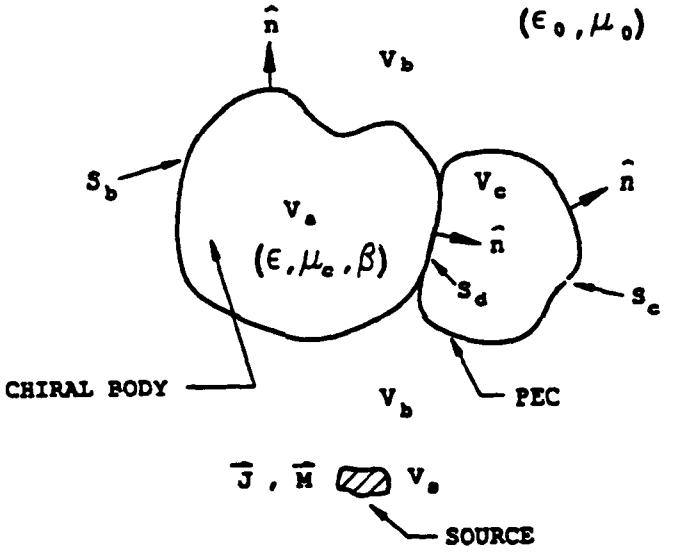


Figure 1. Inhomogeneous chrial body attached to a PEC body.

The first step in the development of the integral equations is to obtain an expression for the fields outside the chiral body. This can be accomplished by means of the vector-dyadic Green's second identity [24], namely

$$\begin{aligned}&\overline{E}^{inc}(\overline{r})+\overline{I}_c(\hat{n}'\times(\nabla'\times\overline{E}_b))\\&+\int_{S_b}\left\{\nabla\times\overline{\overline{G}}_0(\overline{r},\overline{r}')\cdot\hat{n}'\times\overline{E}_b+\hat{n}'\times(\nabla'\times\overline{E}_b)\frac{\psi(R)}{4\pi}\right.\\&\left.-\nabla'_s\cdot(\hat{n}'\times\nabla'\times\overline{E}_b)\frac{\nabla'\psi(R)}{4\pi k_0^2}\right\}ds'\\&=\begin{cases}\overline{E}_b(\overline{r}) & \overline{r}\in V_b\\ 0 & \overline{r}\notin V_b\end{cases}\end{aligned} \quad (8a)$$

where

$$\bar{I}_c(\bar{a}(\bar{r}')) = \int_{S_c} \left\{ \bar{a}(\bar{r}') \frac{\psi(R)}{4\pi} - \nabla'_s \cdot (\bar{a}(\bar{r}')) \frac{\nabla'\psi(R)}{4\pi k_0^2} \right\} ds' \tag{8b}$$

and

$$\nabla'_s = \nabla' - \hat{n}' \frac{\partial}{\partial n'} \tag{8c}$$

The prime on the operator ∇' denotes differentiation with respect to the prime coordinates, while the prime on ds' denotes integration in the prime coordinates. The free space dyadic Green's function $\bar{\bar{G}}_0$ is

$$\bar{\bar{G}}_0(\bar{r},\bar{r}') = \left(\bar{\bar{I}} + \frac{\nabla\nabla}{k_0^2}\right) \frac{\psi(R)}{4\pi} \quad ; \quad \psi(R) = \frac{e^{ik_0 R}}{R} \tag{9a}$$

where $R = |\bar{r} - \bar{r}'|$, $k_0 = \omega\sqrt{\epsilon_0\mu_0}$ (ϵ_0 and μ_0 are the permittivity and permeability of free space, respectively) and

$$\nabla \times \bar{\bar{G}}_0(\bar{r},\bar{r}') = \frac{1}{4\pi} \nabla \times \left(\bar{\bar{I}}\psi(R)\right) \tag{9b}$$

Note that $\bar{\bar{G}}_0$ and $\nabla \times \bar{\bar{G}}_0$ have the following symmetry properties

$$\bar{\bar{G}}_0^T(\bar{r},\bar{r}') = \bar{\bar{G}}_0(\bar{r}',\bar{r}) \quad ; \quad \left(\nabla' \times \bar{\bar{G}}_0(\bar{r}',\bar{r})\right)^T = \nabla \times \bar{\bar{G}}_0(\bar{r},\bar{r}') \tag{9c}$$

where T denotes the transpose operation. The incident field $\bar{E}^{inc}$ in (8a) is given by

$$\bar{E}^{inc}(\bar{r}) = \int_{V_a} \bar{\bar{G}}_0(\bar{r},\bar{r}') \cdot \{i\omega\mu_0 \bar{J}(\bar{r}') - \nabla' \times \bar{M}(\bar{r}')\} dv' \tag{10}$$

It is important to mention that the direct application of the vector-dyadic Green's second identity yields a highly singular integrand $(0(R^{-3}))$ on the surface integrals over S_b and S_c. Since the fields will be evaluated on these surfaces, it is necessary to obtain less singular expressions. The less singular expressions $(0(R^{-2}))$ in (8) were obtained by means of the identity [8]

$$\oint_S \bar{a}(\bar{r}') \cdot \nabla'\nabla'\psi(R) ds' = -\oint_S \nabla'_s \cdot (\bar{a}(\bar{r}')) \nabla'\psi(R) ds' \tag{11}$$

where S is a closed surface and the vector $\bar{a}$ is tangential to the surface S.

The next step in the analysis is to multiply (6a) (right hand side) by $\bar{\bar{G}}_0(\bar{r},\bar{r}')$ and integrate over V_a. By means of the divergence theorem, the following expression can be obtained

$$\begin{aligned}
&-\bar{I}_d\left\{\hat{n}' \times \left(\frac{\tau}{\mu_c}\nabla' \times \bar{E}_a\right)\right\} \\
&-\int_{S_b} \left\{\frac{\tau}{\mu_c}\nabla \times \bar{\bar{G}}_0(\bar{r},\bar{r}') \cdot \hat{n}' \times \bar{E}_b + \left(\hat{n}' \times \frac{\nabla' \times \bar{E}_b}{\mu_0} - \omega^2 \epsilon \beta \hat{n}' \times \bar{E}_b\right) \frac{\psi(R)}{4\pi} \right. \\
&\left. + \left(\omega^2 \nabla'_s \cdot (\epsilon\beta\hat{n}' \times \bar{E}_b) - \frac{\nabla'_s \cdot (\hat{n}' \times (\nabla' \times \bar{E}_b))}{\mu_0}\right) \frac{\nabla'\psi}{4\pi k_0^2} \right\} ds' \\
&-\bar{I}_V(\bar{E}_a(\bar{r}')) = \begin{cases} \bar{E}_a(\bar{r}) \frac{\tau(\bar{r})}{\mu_c(\bar{r})} & \bar{r} \in V_a \\ 0 & \bar{r} \notin V_a \end{cases}
\end{aligned} \tag{12a}$$

where $\overline{\overline{I}}_d(\overline{a}(\overline{r}'))$ is the same as $\overline{\overline{I}}_c(\overline{a}(\overline{r}'))$, except that the integral in (8b) is over S_d instead of S_c. The operator $\overline{\overline{I}}_V$ in (12a) is given by

$$\overline{\overline{I}}_V(\overline{b}(\overline{r}')) = \int_{V_a} \overline{b}(\overline{r}') \cdot \overline{\overline{Q}}(\overline{r}', \overline{r})\, dv' \tag{12b}$$

where $\overline{\overline{Q}}(\overline{r}', \overline{r})$ is a dyadic differential operator, namely

$$\begin{aligned}\overline{\overline{Q}}(\overline{r}', \overline{r}) = &\left\{\nabla'\left(\frac{\tau(\overline{r}')}{\mu_c(\overline{r}')}\right) \times -2\omega^2 \epsilon(\overline{r}')\beta(\overline{r}')\overline{\overline{I}}\cdot\right\} \nabla' \times \overline{\overline{G}}_0(\overline{r}', \overline{r}) \\ &- \omega^2 \nabla'(\epsilon(\overline{r}')\beta(\overline{r}')) \times \overline{\overline{G}}_0(\overline{r}', \overline{r}) + \frac{(\tau(\overline{r}')k_0^2 - k^2(\overline{r}'))}{\mu_c(\overline{r}')} \overline{\overline{G}}_0(\overline{r}', \overline{r})\end{aligned} \tag{12c}$$

The first vector integral equation can be obtained by multiplying (12a) by μ_0 and adding to (8a), namely

$$\begin{aligned}&\overline{\overline{I}}_c(\hat{n}' \times (\nabla' \times \overline{E}_b)) - \mu_0 \overline{\overline{I}}_d\left(\hat{n}' \times \left(\frac{\tau(\overline{r}')}{\mu_c(\overline{r}')} \nabla' \times \overline{E}_a\right)\right) + \overline{\overline{I}}_b(\hat{n}' \times \overline{E}_b) \\ &- \mu_0 \overline{\overline{I}}_V(\overline{E}_a(\overline{r}')) = \begin{cases} \frac{\mu_0 \tau(\overline{r})}{\mu_c(\overline{r})} \overline{E}_a(\overline{r}) - \overline{E}^{inc}(\overline{r}) & \overline{r} \in V_a \\ \overline{E}_b(\overline{r}) - \overline{E}^{inc}(\overline{r}) & \overline{r} \in V_b \\ -\overline{E}^{inc}(\overline{r}) & \overline{r} \in V_c \end{cases}\end{aligned} \tag{13a}$$

where

$$\begin{aligned}\overline{\overline{I}}_b(\overline{a}(\overline{r}')) = \int_{S_b} &\left\{\left[\left(1 - \frac{\mu_0 \tau}{\mu_c}\right) \nabla \times \overline{\overline{G}}_0(\overline{r}, \overline{r}') + \mu_0 \omega^2 \epsilon \beta \frac{\psi(R)}{4\pi} \overline{\overline{I}}\right] \cdot \overline{a}(\overline{r}') \right. \\ &\left. - \omega^2 \mu_0 \nabla'_s \cdot (\epsilon \beta \overline{a}(\overline{r}')) \frac{\nabla' \psi(R)}{4\pi k_0^2}\right\} ds'\end{aligned} \tag{13b}$$

Note that when the medium V_a becomes achiral ($\beta = 0$, $\tau = 1$, $\mu_c = \mu$), $\overline{\overline{Q}}(\overline{r}', \overline{r})$ and $\overline{\overline{I}}_b(\overline{a}(\overline{r}'))$ become

$$\overline{\overline{Q}}(\overline{r}', \overline{r}) = \nabla'\left(\frac{1}{\mu}\right) \times \nabla' \times \overline{\overline{G}}_0(\overline{r}', \overline{r}) + \left(\frac{k_0^2 - k^2}{\mu}\right) \overline{\overline{G}}_0(\overline{r}', \overline{r}) \tag{14a}$$

$$\overline{\overline{I}}_b(\overline{a}(\overline{r}')) = \int_{S_b} \left(1 - \frac{\mu_0}{\mu}\right) \nabla \times \overline{\overline{G}}_0(\overline{r}, \overline{r}') \cdot \overline{a}(\overline{r}')\, ds' \tag{14b}$$

where the parameters ϵ, μ, and k still depend on position, but that dependence is not shown explicitly in (14a) and (14b) to simplify the notation. Furthermore, k becomes the wavenumber of the achiral body V_a when $\beta = 0$.

To solve for the fields $\overline{E}_a$ and $\overline{E}_b$, it follows from (13) that it is necessary to solve for the fields $\overline{E}_a$ in V_a, $\hat{n}' \times \overline{E}_b$ on S_b, $\hat{n}' \times (\nabla' \times \overline{E}_b)$ on S_c, and $\hat{n}' \times \left(\frac{\tau}{\mu_a} \nabla' \times \overline{E}_a\right)$ on S_d, respectively. In other words, there are a total of 9 scalar unknown functions. Equation (13a) where $\overline{r} \in V_a$ yields three scalar equations. Therefore, six additional scalar equations remain to be found. Two

scalar equations are found by taking the cross product of $\hat{n}$ and (13a) and taking the limit of $\bar{r} \in V_b$ as $\bar{r} \to S_c$, namely

$$\hat{n} \times \bar{E}^{inc}(\bar{r}) = -\hat{n} \times P\bar{I}_c\left(\hat{n}' \times (\nabla' \times \bar{E}_b)\right) + \mu_0 \hat{n} \times \bar{I}_d\left(\hat{n}' \times \left(\frac{\tau}{\mu_c}\nabla' \times \bar{E}_a\right)\right) - \hat{n} \times \bar{I}_b\left(\hat{n}' \times \bar{E}_b\right) + \mu_0 \hat{n} \times \bar{I}_V\left(\bar{E}_a(\bar{r}')\right) \tag{15a}$$

where P in front of the operator $\bar{I}_c$ denotes the principal value. Similar equations can be obtained by following the same procedure as above. Thus, for $\bar{r} \in V_b$ when $\bar{r} \to S_b$

$$-\hat{n} \times \bar{E}^{inc}(\bar{r}) + \frac{1}{2}\left(1 + \mu_0 \frac{\tau}{\mu_c}\right)\hat{n} \times \bar{E}_b = \hat{n} \times \bar{I}_c\left(\hat{n}' \times (\nabla' \times \bar{E}_b)\right) - \mu_0 \hat{n} \times \bar{I}_d\left(\hat{n}' \times \left(\frac{\tau}{\mu_c}\nabla' \times \bar{E}_a\right)\right) + \hat{n} \times P\bar{I}_b\left(\hat{n}' \times \bar{E}_b\right) - \mu_0 \hat{n} \times \bar{I}_V\left(\bar{E}_a(\bar{r}')\right) \tag{15b}$$

and for $\bar{r} \in V_a$ when $\bar{r} \to S_d$

$$\hat{n} \times \bar{E}^{inc}(\bar{r}) = -\hat{n} \times \bar{I}_c\left(\hat{n}' \times (\nabla' \times \bar{E}_b)\right) + \mu_0 \hat{n} \times P\bar{I}_d\left(\hat{n}' \times \left(\frac{\tau}{\mu_c}\nabla' \times \bar{E}_a\right)\right) - \hat{n} \times \bar{I}_b\left(\hat{n}' \times \bar{E}_b\right) + \mu_0 \hat{n} \times \bar{I}_V\left(\bar{E}_a(\bar{r}')\right) \tag{15c}$$

As was mentioned before, the integral equations obtained in this section can also be developed by starting with the equivalent polarization current formulation. That procedure shows that the equivalent polarization current method yields expressions where the electric and magnetic fields do indeed satisfy the correct boundary conditions. This fact is not always clear because no boundary conditions are explicitly imposed in deriving the equivalent polarization currents.

III. MOMENT METHOD SOLUTION

In general, the integral equations developed in the previous section cannot be solved in closed form, except for special cases as shown in Section V. Thus, it becomes necessary to use some type of numerical technique to solve these equations. There are several numerical techniques to solve integral equations like the ones developed in the previous section. In this section, a brief outline of a moment method solution will be given [6,25]. The first step is to expand the unknown functions in terms of a set of basis functions. That is, $\bar{E}_a(\bar{r}')$ can be expressed as

$$\bar{E}_a(\bar{r}') = \sum_{\ell=1}^{3} \hat{x}_\ell \bar{E}_a(\bar{r}') \cdot \hat{x}_\ell \quad ; \quad \hat{x}_\ell \cdot \bar{E}_a(\bar{r}') = \sum_{n=1}^{N} \alpha_{n'}^a f_{n'}^a(\bar{r}') \quad ; \quad \bar{r}' \in V_a \tag{16a}$$

where $n' = n + (\ell - 1)N$ and $\hat{x}_\ell (\ell = 1, 2, 3)$ is an orthonormal set of coordinates in V_a and $\{f_{n'}^a\}$ is a set of basis functions defined in V_a. Likewise, the other unknown functions can also be expanded in terms of basis functions defined in the domains of S_b, S_c, and S_d, respectively. That is,

$$\hat{n}' \times \overline{E}_b(\overline{r}') = \sum_{\ell=1}^{2} \hat{t}_\ell^b \sum_{n=1}^{M_b} \alpha_{n''}^b f_{n''}^b(\overline{r}') \ ; \ \overline{r}' \in S_b \text{ and } n'' = n + (\ell-1)M_b \tag{16b}$$

$$\hat{n}' \times \nabla' \times \overline{E}_b(\overline{r}') = \sum_{\ell=1}^{2} \hat{t}_\ell^c \sum_{m=1}^{M_c} \alpha_{m'}^c f_{m'}^c(\overline{r}') \ ;$$
$$\overline{r}' \in S_c \text{ and } m' = m + (\ell-1)M_c \tag{16c}$$

and

$$\frac{\tau}{\mu_c} \hat{n}' \times \nabla' \times \overline{E}_a(\overline{r}') = \sum_{\ell=1}^{2} \hat{t}_\ell^d \sum_{m=1}^{M_d} \alpha_{m''}^d f_{m''}^d(\overline{r}') \ ;$$
$$\overline{r}' \in S_d \text{ and } m'' = m + (\ell-1)M_d \tag{16d}$$

where $\{\hat{t}_\ell^\rho\}_{\ell=1}^2 (\rho = b, c, d)$ is an orthonormal set defined on $S_\rho (\rho = b, c, d)$. Note that the basis functions defined in (16b)–(16d) have to be differentiable. Substituting (16) into (13a), one gets

$$\sum_{\ell=1}^{2} \left\{ \sum_{n=1}^{M_b} \alpha_{n''}^b \overline{I}_b \left(\hat{t}_\ell^b f_{n''}^b(\overline{r}') \right) + \sum_{m=1}^{M_c} \alpha_{m'}^c \overline{I}_c \left(\hat{t}_\ell^c f_{m'}^c(\overline{r}') \right) \right.$$
$$\left. - \mu_0 \sum_{m=1}^{M_d} \alpha_{m''}^d \overline{I}_d \left(\hat{t}_\ell^d f_{m''}^d(\overline{r}') \right) \right\}$$
$$- \mu_0 \sum_{\ell=1}^{3} \left\{ \sum_{n=1}^{N} \alpha_{n'}^a \left[\overline{I}_V \left(\hat{x}_\ell f_{n'}^a(\overline{r}') \right) + \frac{\tau(\overline{r})}{\mu_c(\overline{r})} \hat{x}_\ell f_{n'}^a(\overline{r}) \right] \right\} = -\overline{E}^{inc}(\overline{r}) \ ; \ \overline{r} \in V_a \tag{17}$$

Similar equations can be obtained by substituting (16) into (15). The next step in the moment method procedure is to define some testing function in the domains of V_a, S_b, S_c, and S_d. Let

$$\overline{w}^a(\overline{r}) = \sum_{\ell=1}^{3} \hat{x}_\ell \sum_{n=1}^{N} w_{n'}^a(\overline{r}) \ , \ \overline{r} \in V_a \tag{18a}$$

where $n' = n + (\ell-1)N$ and

$$\overline{w}^\rho(\overline{r}) = \sum_{\ell=1}^{2} \hat{t}_\ell^\rho \sum_{n=1}^{M_\rho} w_{n''}^\rho(\overline{r}) \ , \ \overline{r} \in S_\rho \qquad (\rho = b, c, d) \tag{18b}$$

where $n'' = n + (\ell-1)M_\rho$. Next, let's introduce an inner product defined as follows

$$\langle \overline{a}(\overline{r}), \overline{b}(\overline{r}) \rangle_{V_a} = \int_{V_a} \overline{a}(\overline{r}) \cdot \overline{b}(\overline{r}) dv \tag{19a}$$

$$\langle \overline{a}(\overline{r}), \overline{b}(\overline{r}) \rangle_{S_\rho} = \int_{S_\rho} \overline{a}(\overline{r}) \cdot \overline{b}(\overline{r}) ds \qquad \rho = b, c, d \tag{19b}$$

The original integral equations of Section II can then be transformed into a linear system of simultaneous equations by substituting (16) into (15) and incorporating (19) into these resulting expressions and into (17), namely

$$[V] = [Z][I] \tag{20}$$

where

$$[I] = \left[\alpha_1^a, \ldots, \alpha_{3N}^a, \alpha_1^b, \ldots, \alpha_{2M_b}^b, \alpha_1^c, \ldots, \alpha_{2M_c}^c, \alpha_1^d, \ldots, \alpha_{2M_d}^d\right]^T \tag{21a}$$

and

$$[V] = \left[V_1^a, \ldots, V_{3N}^a, V_1^b, \ldots, V_{2M_b}^b, V_1^c, \ldots, V_{2M_c}^c, V_1^d, \ldots, V_{2M_d}^d\right]^T \tag{21b}$$

are $(3N + 2(M_b + M_c + M_d)) \times 1$ column vectors and $[Z]$ is a square matrix. The V_n^ρ elements are defined as follows

$$V_{n'}^a = \left\langle \overline{E}^{inc}(\overline{r}), \hat{x}_\ell w_{n'}^a(\overline{r}) \right\rangle_{V_a} \quad ; \quad 1 \le n \le N,\ 1 \le \ell \le 3 \tag{22a}$$

where $n' = n + (\ell - 1)N$ and

$$V_{n'}^\rho = \left\langle \overline{E}^{inc}(\overline{r}), \hat{t}_\ell^\rho w_{n'}^\rho(\overline{r}) \right\rangle_{S_\rho} \quad ; \quad 1 \le n \le M_\rho,\ 1 \le \ell \le 2,\ \rho = b, c, d \tag{22b}$$

where $n' = n + (\ell - 1)M_\rho$. The matrix $[Z]$, which is sometimes referred to as the impedance matrix, can be expressed as follows:

$$[Z] = \begin{array}{c} \begin{array}{cccc} \leftarrow 3N \rightarrow & \leftarrow 2M_b \rightarrow & \leftarrow 2M_c \rightarrow & \leftarrow 2M_d \rightarrow \end{array} \\ \left[\begin{array}{c:c:c:c} [Z_{nm}^{aa}] & [Z_{nm}^{ab}] & [Z_{nm}^{ac}] & [Z_{nm}^{ad}] \\ \hdashline [Z_{nm}^{ba}] & [Z_{nm}^{bb}] & [Z_{nm}^{bc}] & [Z_{nm}^{bd}] \\ \hdashline [Z_{nm}^{ca}] & [Z_{nm}^{cb}] & [Z_{nm}^{cc}] & [Z_{nm}^{cd}] \\ \hdashline [Z_{nm}^{da}] & [Z_{nm}^{db}] & [Z_{nm}^{dc}] & [Z_{nm}^{dd}] \end{array}\right] \end{array} \begin{array}{l} \uparrow \\ 3N \\ \downarrow \\ \uparrow \\ 2M_b \\ \downarrow \\ \uparrow \\ 2M_c \\ \downarrow \\ \uparrow \\ 2M_d \\ \downarrow \end{array} \tag{23}$$

Note that the sub-matrices $Z_{nm}^{\rho\rho}(\rho = a, b, c, d)$ are square matrices, whereas the rest of the sub-matrices in (23) are rectangular. All the matrix elements in (23) are given in the Appendix. The solution of (20) yields the values of the expansion coefficients introduced in (16).

IV. FAR-ZONE SCATTERED FIELD

Once the integral equations developed in Section II are solved either by the moment method or any other technique, it is of interest to calculate the scattered field in the far-zone. To accomplish this task it is necessary to first obtain expressions of $0\left(R_0^{-1}\right)$ for the Green's functions $\overline{\overline{G}}_0\left(\overline{r},\overline{r}'\right)$ and $\nabla\times\overline{\overline{G}}_0\left(\overline{r},\overline{r}'\right)$, where R_0 is depicted in Fig. 2. As it is well known, in the far zone, the vectors $\overline{r}$ and $\overline{r}-\overline{r}'$ are assumed to be parallel and R^{-1} is assumed to be equal to R_0^{-1} in the denominator of $\psi(R)$. Thus, keeping these assumptions in mind, one gets

$$\overline{\overline{G}}_0\left(\overline{r},\overline{r}'\right)=\overline{\overline{G}}_0\left(\overline{r}',\overline{r}\right)\sim\left(\widehat{\phi}\widehat{\phi}+\widehat{\theta}\widehat{\theta}\right)e^{-ik_0\overline{r}'\cdot\hat{r}}\frac{\psi(R_0)}{4\pi} \tag{24a}$$

$$\nabla\times\overline{\overline{G}}_0\left(\overline{r},\overline{r}'\right)=-\nabla'\times\overline{\overline{G}}_0\left(\overline{r}',\overline{r}\right)\sim ik_0\left(\widehat{\phi}\widehat{\theta}-\widehat{\theta}\widehat{\phi}\right)e^{-ik_0\overline{r}'\cdot\hat{r}}\frac{\psi(R_0)}{4\pi} \tag{24b}$$

where $\hat{\phi}$ and $\hat{\theta}$ are the usual unit vectors in the unprimed spherical coordinate system transverse to the radial unit vector $\hat{r}$. Substituting (24a) and (24b) into (13a) for $\overline{r}\in V_b$ yields the following far-zone expression

$$\overline{E}_b^s(\overline{r})=\overline{E}_b(\overline{r})-\overline{E}^{inc}(\overline{r})\sim\overline{F}(\theta,\phi)\frac{\psi(R_0)}{4\pi k_0}\quad;\quad\begin{matrix}\overline{r}\in V_b\\ |\overline{r}|>>|\overline{r}'|\end{matrix} \tag{25a}$$

If the moment method procedure developed in the previous section is followed to solve the integral equations, it yields the following expression for $\overline{F}(\theta,\phi)$

$$\begin{aligned}\overline{F}(\theta,\phi)=&-\mu_0k_0\sum_{\ell=1}^{3}\sum_{n=1}^{N}\alpha_{n'}^{a}\overline{I}_V^f\left(\hat{x}_\ell f_{n'}^{a}\left(\overline{r}'\right)\right)+k_0\sum_{\ell=1}^{2}\left\{\sum_{n=1}^{M_b}\alpha_{n''}^{b}\overline{I}_b^f\left(\hat{t}_\ell^{b}f_{n''}^{b}\left(\overline{r}'\right)\right)\right.\\&\left.+\sum_{m=1}^{M_c}\alpha_{m'}^{c}\overline{I}_c^f\left(\hat{t}_\ell^{c}f_{m'}^{c}\left(\overline{r}'\right)\right)-\mu_0\sum_{m=1}^{M_d}\alpha_{m''}^{d}\overline{I}_d^f\left(\hat{t}_\ell^{d}f_{m''}^{d}\left(\overline{r}'\right)\right)\right\}\end{aligned} \tag{25b}$$

where n', n'', m', and m'' are defined in (16). The integral operators $\overline{I}^f$ are given by

$$\begin{aligned}\overline{I}_b^f\left(\overline{a}\left(\overline{r}'\right)\right)=&ik_0(\widehat{\phi}\widehat{\theta}-\widehat{\theta}\widehat{\phi})\cdot\int_{S_b}\left(1-\frac{\mu_0\tau}{\mu_c}\right)e^{-ik_0\overline{r}'\cdot\hat{r}}\overline{a}\left(\overline{r}'\right)ds'\\&+\mu_0\omega^2(\widehat{\phi}\widehat{\phi}+\widehat{\theta}\widehat{\theta})\cdot\int_{S_b}\epsilon\beta e^{-ik_0\overline{r}'\cdot\hat{r}}\overline{a}\left(\overline{r}'\right)ds'\end{aligned} \tag{26a}$$

$$\overline{I}_\rho^f\left(\overline{a}\left(\overline{r}'\right)\right)=(\widehat{\phi}\widehat{\phi}+\widehat{\theta}\widehat{\theta})\cdot\int_{S_\rho}\overline{a}\left(\overline{r}'\right)e^{-ik_0\overline{r}'\cdot\hat{r}}ds'\quad;\quad\rho=c,d \tag{26b}$$

$$\begin{aligned}\overline{I}_V^f\left(\overline{b}\left(\overline{r}'\right)\right)=&-ik_0\int_{V_a}e^{-ik_0\overline{r}'\cdot\hat{r}}\overline{b}\left(\overline{r}'\right)\\&\cdot\left\{\nabla'\left(\frac{\tau}{\mu_c}\right)\times-2\omega^2\epsilon\beta\overline{\overline{I}}\cdot\right\}\left(-\widehat{\theta}\widehat{\phi}+\widehat{\phi}\widehat{\theta}\right)dv'\\&+\omega^2\int_{V_a}e^{-ik_0\overline{r}'\cdot\hat{r}}\overline{b}\left(\overline{r}'\right)\end{aligned}$$

$$\cdot \left\{ -\nabla'(\epsilon\beta) \times + (\tau\mu_0\epsilon_0 - \mu_c\epsilon) \frac{\overline{\overline{I}}}{\mu_c} \cdot \right\} \left(\widetilde{\phi\phi} + \widetilde{\theta\theta} \right) dv' \qquad (26c)$$

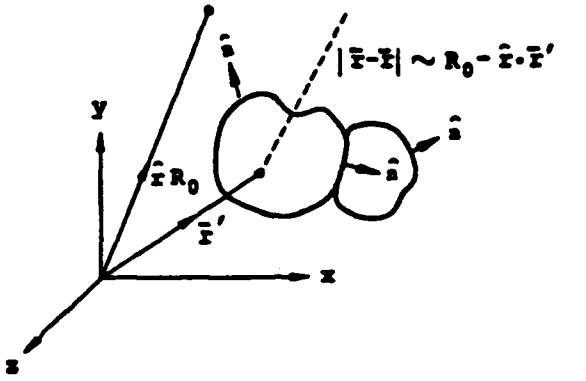


Figure 2. Definition of R_0 and $\bar{r} - \bar{r}'$ in far field approximation.

V. FIELDS INDUCED IN AND SCATTERED BY AN ELECTRICALLY SMALL CHIRAL SPHERE

First, let's consider the fields induced in a homogeneous chiral sphere of radius a, where a is much smaller than the two wavelengths inside the material, such that the fields within the sphere are constant. Without loss of generality and to simplify the analysis, the center of the sphere is assumed to be located at the origin of the coordinate system. Thus, evaluating the fields at $\bar{r} = \bar{0}$, (13a) can be written as follows:

$$\bar{I}_b (\hat{n}' \times \bar{E}_a) - \mu_0 \bar{I}_V (\bar{E}_a) = \frac{\mu_0 \tau}{\mu_c} \bar{E}_a - \bar{E}^{inc} \quad ; \quad \bar{r} = \bar{0} \qquad (27)$$

Since $\bar{E}_a$ and $\bar{H}_a$ are constant inside the chiral sphere, the evaluation of $\bar{I}_b$ and I_V yields the following closed forms expressions

$$\bar{I}_b = \left(-\frac{2}{3} + \frac{2\mu_0\tau}{3\mu_c} + \frac{\epsilon}{3\epsilon_0} \left(1 - \frac{1}{\tau} \right) \right) \bar{E}_a - \frac{ik^2\beta\eta_0}{3k_0\tau} \bar{H}_a \qquad (28a)$$

$$\bar{I}_V = -\frac{\bar{E}_a}{3\mu_c k_0^2} \left(\tau k_0^2 - k^2 \right) \qquad (28b)$$

where $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$. Because the expressions for $\bar{I}_b$ and $\bar{I}_v$ contain two unknowns, namely, the magnetic $\bar{H}_a$ and electric $\bar{E}_a$ fields, (27) is not sufficient to determine both unknowns and therefore a second equation is needed. Another equation is obtained by taking the curl of (27) and evaluating the resulting expression at $\bar{r} = \bar{0}$, namely,

$$\nabla \times \bar{I}_b - \mu_0 \nabla \times \bar{I}_V = \frac{\mu_0 \tau}{\mu_c} \left(\frac{i\omega\mu_c}{\tau} \bar{H}_a + \frac{k^2\beta}{\tau} \bar{E}_a \right) - \nabla \times \bar{E}^{inc} \quad ; \quad \bar{r} = \bar{0} \qquad (29)$$

Once again, $\nabla \times \bar{I}_b$ and $\nabla \times \bar{I}_V$ can be evaluated in closed form because of the

assumption that the fields are constant within the chiral sphere. Thus, at $\bar{r} = \bar{0}$

$$\nabla \times \bar{J}_b = -\frac{1}{3}\left(1 - \frac{\mu_0 \tau}{\mu_c}\right)\left(\frac{i\omega\mu_c \bar{H}_a}{\tau} + \frac{k^2\beta}{\tau}\bar{E}_a\right) - \frac{2}{3}\mu_0\omega^2\epsilon\beta\bar{E}_a \quad (30a)$$

$$\nabla \times \bar{J}_V = -\frac{4}{3}\omega^2\epsilon\beta\bar{E}_a \quad (30b)$$

The fields $\bar{E}_a$ and $\bar{H}_a$ can now be obtained by solving the two simultaneous equations in (27) and (29). Thus, substituting (30) into (29) and (28) into (27), respectively, and solving (27) and (29), yields

$$\bar{E}_a = \left\{3\left(2 + \frac{\mu_{cr}}{\tau}\right)\bar{E}^{inc}(\bar{r} = \bar{0}) - 3\left(\frac{\epsilon_r\mu_{cr}k_0\beta}{\tau}\right)i\eta_0\bar{H}^{inc}(\bar{r} = \bar{0})\right\}\zeta^{-1} \quad (31a)$$

$$\bar{H}_a = \left\{3\left(2 + \frac{\epsilon_r}{\tau}\right)\bar{H}^{inc}(\bar{r} = \bar{0}) + i3\left(\frac{\epsilon_r\mu_{cr}k_0\beta}{\tau}\right)\frac{\bar{E}^{inc}(\bar{r} = \bar{0})}{\eta_0}\right\}\zeta^{-1} \quad (31b)$$

where $\nabla \times \bar{E}^{inc} = ik_0\eta_0\bar{H}^{inc}$, $\epsilon_r = \epsilon/\epsilon_0$, $\mu_{cr} = \mu_c/\mu_0$, and

$$\zeta = \left(2 + \frac{\mu_{cr}}{\tau}\right)\left(2 + \frac{\epsilon_r}{\tau}\right) - \left(\frac{\epsilon_r\mu_{cr}k_0\beta}{\tau}\right)^2 \quad (31c)$$

As expected, each of the expressions for $\bar{E}_a$ and $\bar{H}_a$ contain cross-polarized components due to the presence of the chiral parameter β. When the sphere becomes achiral ($\beta = 0$), (31) reduces to a well known result, namely

$$\bar{E}_a = \frac{3}{2 + \epsilon_r}\bar{E}^{inc}(\bar{r} = \bar{0}) \quad ; \quad \bar{H}_a = \frac{3}{2 + \mu_r}\bar{H}^{inc}(\bar{r} = \bar{0}) \quad (32)$$

where $\mu_r = \mu/\mu_0$.

The results given in (31) are the same as those obtained by Lindell and Sihvola [26] who followed a totally different approach to obtain the fields inside the chiral sphere. The induced magnetic $(\bar{P}_m)$ and electric $(\bar{P}_e)$ dipole moments can also be computed. However, it is necessary to first compute the equivalent magnetic $(\bar{M}_{eq})$ and electric $(\bar{J}_{eq})$ polarization currents [6]. These currents, which are assumed to be radiating in free space, are given by

$$\bar{J}_{eq} = \begin{cases} k_0\left[-\frac{i}{\eta_0}\left(\frac{\epsilon_r}{\tau} - 1\right)\bar{E}_a + \frac{k_0\beta}{\tau}\epsilon_r\mu_{cr}\bar{H}_a\right] & ; \bar{r} \in V_a \\ 0 & ; \bar{r} \notin V_a \end{cases} \quad (33a)$$

$$\bar{M}_{eq} = \begin{cases} k_0\left[-i\eta_0\left(\frac{\mu_{cr}}{\tau} - 1\right)\bar{H}_a - \frac{k_0\beta}{\tau}\epsilon_r\mu_{cr}\bar{E}_a\right] & ; \bar{r} \in V_a \\ 0 & ; \bar{r} \notin V_a \end{cases} \quad (33b)$$

where V_a is the volume occupied by the sphere.

Since $\bar{E}_a$ and $\bar{H}_a$ are constant, $\bar{P}_e$ and $\bar{P}_m$ can be easily computed, namely,

$$\bar{P}_e = \bar{J}_{eq}\,4\pi a^3/3 \quad ; \quad \bar{P}_m = \bar{M}_{eq}\,4\pi a^3/3 \quad (34)$$

Substituting (31) and (33) into (34), one gets

$$\bar{P}_e = \frac{4\pi a^3 k_0}{\zeta\tau}\left\{3\epsilon_r\mu_{cr}k_0\beta\bar{H}^{inc}(0)\right.$$

$$+\frac{i}{\eta_0}\overline{E}^{inc}(0)\left[(2+\mu_{cr})(1-\epsilon_r)-2(k_0\beta)^2\epsilon_r\mu_{cr}\right]\Big\} \tag{35a}$$

$$\overline{P}_m=\frac{4\pi a^3k_0}{\zeta_r}\Big\{-3\epsilon_r\mu_{cr}k_0\beta\overline{E}^{inc}(0)$$
$$+i\eta_0\overline{H}^{inc}(0)\left[(2+\epsilon_{cr})(1-\mu_r)-2(k_0\beta)^2\epsilon_r\mu_{cr}\right]\Big\} \tag{35b}$$

Note that in a chiral scatterer, the electric and magnetic dipole moments depend on both the incident electric and magnetic fields. For an achiral body ($\beta = 0$), the expressions for $\overline{P}_{e,m}$ are much simpler, namely

$$\overline{P}_e=4\pi a^3k_0\frac{i}{\eta_0}\overline{E}^{inc}(0)\left(\frac{1-\epsilon_r}{2+\epsilon_r}\right)\quad;\quad\overline{P}_m=4\pi a^3k_0i\eta_0\overline{H}^{inc}(0)\left(\frac{1-\mu_r}{2+\mu_r}\right) \tag{36}$$

where $\overline{P}_e$ is proportional to $\overline{E}^{inc}$ and $\overline{P}_m$ is proportional to $\overline{H}^{inc}$.

Once the dipole moments are obtained, there is a straightforward procedure to calculate the far-zone scattered fields if it is assumed that $\overline{P}_{e,m}$ are infinitesmal dipole moments located at the origin $\overline{r}=\overline{0}$ [27]. The far-zone pattern vector $\overline{F}(\theta,\phi)$, introduced in (25*a*), is given by

$$\overline{F}(\theta,\phi)=\hat{\theta}F_\theta(\theta,\phi)+\hat{\phi}F_\phi(\theta,\phi)=ik_0^2\eta_0\left(\widehat{\theta\theta}+\widehat{\phi\phi}\right)\cdot\overline{P}_e+ik_0^2\left(\widehat{\theta\phi}-\widehat{\phi\theta}\right)\cdot\overline{P}_m \tag{37}$$

where $\overline{P}_e$ and $\overline{P}_m$ are given in (35*a*) and (35*b*), respectively. For $\beta = 0$, the expression for $\overline{F}(\theta,\phi)$ becomes

$$\overline{F}(\theta,\phi)=-4\pi(k_0a)^3\Big\{\left(\widehat{\theta\theta}+\widehat{\phi\phi}\right)\cdot\overline{E}^{inc}(0)\left(\frac{1-\epsilon_r}{2+\epsilon_r}\right)$$
$$+\eta_0\left(\widehat{\theta\phi}-\widehat{\phi\theta}\right)\cdot\overline{H}^{inc}(0)\left(\frac{1-\mu_r}{2+\mu_r}\right)\Big\} \tag{38}$$

Another parameter of interest is the echo area (or radar cross section) $\sigma(\theta,\phi)$, which is defined as follows

$$\sigma(\theta,\phi)=\lim_{R_0\to\infty}4\pi R_0^2\frac{\left|\overline{E}^s\right|^2}{\left|\overline{E}^{inc}(0)\right|^2}=\frac{\left|\overline{F}(\theta,\phi)\right|^2}{4\pi k_0^2\left|\overline{E}^{inc}(0)\right|^2} \tag{39a}$$

Substituting (25*a*) and (37) into (39*a*) yields

$$\frac{\sigma(\theta,\phi)}{\pi a^2}=\frac{k_0^2\left|\eta_0\hat{\theta}\cdot\overline{P}_e+\hat{\phi}\cdot\overline{P}_m\right|^2+k_0^2\left|\eta_0\hat{\phi}\cdot\overline{P}_e-\hat{\theta}\cdot\overline{P}_m\right|^2}{4(\pi a)^2\left|\overline{E}^{inc}(0)\right|^2} \tag{39b}$$

Keeping in mind that $\overline{P}_{e,m}$ are proportional to k_0a^3, it is clear that the expression in (39*b*) satisfies the Rayleigh Law of scattering. That is, for a small chiral sphere, $\sigma(\theta,\phi)/(\pi a^2)$ is proportional to $(k_0a)^4$. As was the case before, (39*b*) becomes much simpler when $\beta = 0$

$$\frac{\sigma(\theta,\phi)}{\pi a^2}=\frac{4(k_0a)^4}{\left|\overline{E}^{inc}(0)\right|^2}\left\{\left|\hat{\theta}\cdot\overline{E}^{inc}(0)\left(\frac{1-\epsilon_r}{2+\epsilon_r}\right)+\eta_0\hat{\phi}\cdot\overline{H}^{inc}(0)\left(\frac{1-\mu_r}{2+\mu_r}\right)\right|^2\right.$$

$$+\left|\hat{\phi}\cdot\overline{E}^{inc}(0)\left(\frac{1-\epsilon_r}{2+\epsilon_r}\right)-\eta_0\hat{\theta}\cdot\overline{H}^{inc}(0)\left(\frac{1-\mu_r}{2+\mu_r}\right)\right|^2\Bigg\} \tag{39c}$$

For purposes of illustration, assume that the incident field is a plane wave field given by

$$\overline{E}^{inc}(\overline{r})=\hat{x}E_0e^{ik_0z}\quad;\quad\overline{H}^{inc}(\overline{r})=\hat{y}\frac{E_0}{\eta_0}e^{ik_0z} \tag{40}$$

where E_0 is a constant. For this incident field and for a sphere of radius $0.005\,\lambda_0$, where λ_0 is the free space wavelength, F_θ, F_ϕ, and the bistatic echo area $\sigma(\theta,\phi)/(\pi a^2)$ are calculated in the E- and H-planes. That is, in Fig. 3 the E-plane patterns are shown, where $\phi=0$ and $0<\theta<\pi$, and in Fig. 4 the H-plane patterns are depicted where $\phi=\pi/2$ and $0<\theta<\pi$. For the parameters given in Figs. 3 and 4, the wavelengths inside the chiral material are $0.5036\,\lambda_0$ and $0.2036\,\lambda_0$. Note that for the achiral case ($\beta=0$), the component F_θ is zero in the H-plane, whereas, the F_ϕ component is zero in the E-plane. In many applications, it is important to know the backscattered echo area. Thus, for $\theta=\pi$,

$$\frac{\sigma(\theta=\pi,\phi)}{\pi a^2}=\frac{36(k_0a)^4\,|\mu_{cr}-\epsilon_r|^2}{\left|(2+\epsilon_r)(2+\mu_{cr})-4\epsilon_r\mu_{cr}(k_0\beta)^2\right|^2} \tag{41}$$

An interesting result is that the backscattered echo area of a small chiral (or achiral) sphere is zero if $\epsilon_r=\mu_{cr}$.

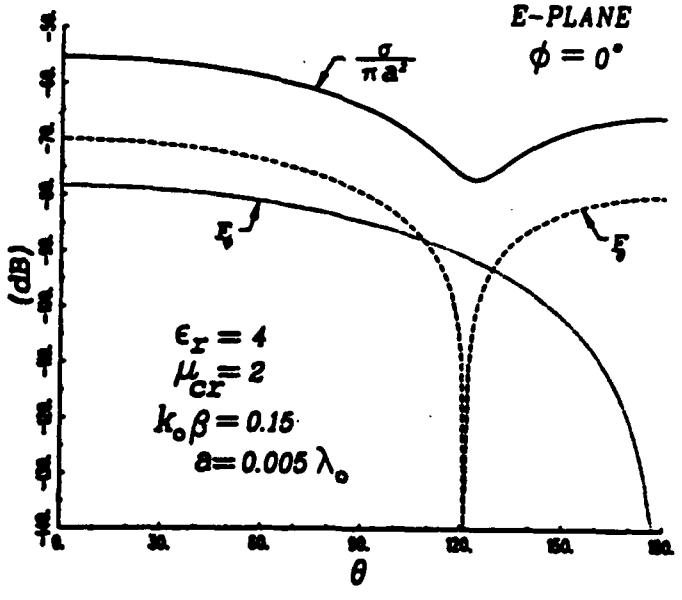


Figure 3. E-plane patterns.

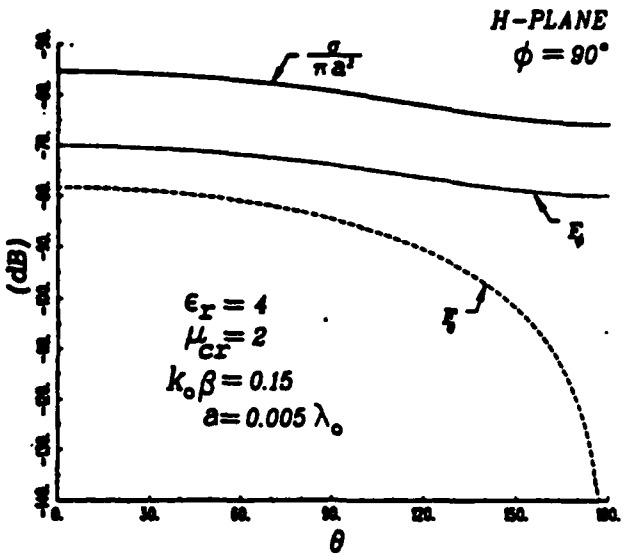


Figure 4. H-plane patterns.

VI. CONCLUSIONS

A set of integral equations for the electric field has been developed for the scattering by an inhomogeneous three dimensional chiral body in the presence of a perfect conductor. Note that a similar set of equations can also be developed for the magnetic field instead of the electric field. The integral equations are obtained by starting with the vector differential wave equations satisfied by the electric field inside and outside the chiral body. These equations are then transformed into integral equations by means of the vector-dyadic Green's second identity, the boundary conditions satisfied by the fields and various other vector-dyadic identities. The integral equations are expressed in terms of a volume integral where the unknown is the electric field inside the chiral body, and surface integrals where the unknowns are the tangential components of the electric field along the chiral-free space boundary and the tangential component of the curl of the electric field along the surface of the perfect conductor. These integral equations are numerically more efficient than the equations obtained in [7], especially when the body is electrically large.

In general, the integral equations developed here cannot be solved in closed form. Thus, one is forced to use some type of numerical procedure. The well known moment method technique is used here to transform the integral equations into a system of linear simultaneous equations. Expressions are obtained here for the so-called impedance matrix in terms of an arbitrary set of basis and testing functions. Since one of the applications of the present work is the study of the scattering properties of chiral bodies, far-zone expressions are obtained for the electric field scattered by a chiral body. Finally, to verify the validity of the in-

tegral equations developed here, they are used to obtain closed form expressions for the fields and dipole moments induced inside an electrically small homogeneous chiral sphere. The dipole moments are then used to calculate the far-zone scattered fields and the radar cross section of the small sphere. As expected, the results obtained for the electrically small chiral sphere show the presence of cross-polarized components of the electric and magnetic fields induced in a chiral body. It is also shown that the echo area satisfies the Rayleigh Law of scattering.

APPENDIX

The elements of the impedance matrix introduced in (23) are given in this appendix. First, the elements of the square sub-matrices are given.

$$Z_{nm}^{aa} = \mu_0 \left\langle \frac{\tau(\overline{r})}{\mu_c(\overline{r})} \widehat{x}_\ell f_m^a(\overline{r}) + \overline{I}_V \left(\widehat{x}_\ell f_m^a(\overline{r})\right) \, , \, \widehat{x}_i \omega_n^a(\overline{r}) \right\rangle_{V_a} \tag{42a}$$

where

$$m = q + (\ell - 1)N \, , \; n = r + (i-1)N \, ; \; 1 \le q \, , \, r \le N \, , \; 1 \le i \, , \; \ell \le 3$$

$$Z_{nm}^{bb} = \left\langle \frac{1}{2}\left(1 + \mu_0 \frac{\tau}{\mu_c}\right) \widehat{t}_\ell^b f_m^b(\overline{r}) - \widehat{n} \times P\overline{I}_b \left(\widehat{t}_\ell^b f_m^b(\overline{r}')\right) \, , \, \widehat{t}_i^b \omega_n^b(\overline{r}) \right\rangle_{S_b} \tag{42b}$$

$$Z_{nm}^{cc} = \left\langle - \widehat{n} \times P\overline{I}_c \left(\widehat{t}_\ell^c f_m^c(\overline{r}')\right) \, , \, \widehat{t}_i^c \omega_n^c(\overline{r}) \right\rangle_{S_c} \tag{42c}$$

$$Z_{nm}^{dd} = \mu_0 \left\langle \widehat{n} \times P\overline{I}_d \left(\widehat{t}_\ell^d f_m^d(\overline{r}')\right) \, , \, \widehat{t}_i^d \omega_n^d(\overline{r}) \right\rangle_{S_d} \tag{42d}$$

In (42b)-(42d), $m = q + (\ell - 1)M_\rho$, $n = r + (i-1)M_\rho$ where $1 \le q, \, r \le M_\rho$, and $1 \le i, \, \ell \le 2$. The subscript ρ in M_ρ is equal to b in (42*b*), c in (42*c*) and d in (42*d*), respectively. All the matrices above are square. The rest of the matrices given below are rectangular.

$$Z_{nm}^{a\rho} = -\left\langle \overline{I}_\rho \left(\widehat{t}_\ell^\rho f_m^\rho(\overline{r}')\right) \, , \, \widehat{x}_i \omega_n^a(\overline{r}) \right\rangle_{V_a} \quad ; \quad \rho = b, c \tag{43a}$$

$$Z_{nm}^{ad} = \mu_0 \left\langle \overline{I}_d \left(\widehat{t}_\ell^d f_m^d(\overline{r}')\right) \, , \, \widehat{x}_i \omega_n^a(\overline{r}) \right\rangle_{V_a} \tag{43b}$$

$$Z_{mn}^{\rho a} = \mu_0 \left\langle \widehat{n} \times \overline{I}_V \left(\widehat{x}_i f_n^a(\overline{r}')\right) \, , \, \widehat{t}_\ell^\rho \omega_m^\rho(\overline{r}) \right\rangle_{S_\rho} \quad ; \quad \rho = b, c, d \tag{43c}$$

where

$$m = q + (\ell - 1)M_\rho \, ; \; 1 \le q \le M_\rho \, , \; 1 \le \ell \le 2$$
$$n = r + (i-1)N \, ; \; 1 \le r \le N \, , \; 1 \le i \le 3$$

Furthermore,

$$Z_{nm}^{bc} = \left\langle - \widehat{n} \times \overline{I}_c \left(\widehat{t}_\ell^c f_m^c(\overline{r}')\right) \, , \, \widehat{t}_i^b \omega_n^b(\overline{r}) \right\rangle_{S_b} \tag{44a}$$

$$Z_{nm}^{bd} = \mu_0 \left\langle \widehat{n} \times \overline{I}_d \left(\widehat{t}_\ell^d f_m^d(\overline{r}')\right) \, , \, \widehat{t}_i^b \omega_n^b(\overline{r}) \right\rangle_{S_b} \tag{44b}$$

$$Z_{mn}^{\rho b} = \left\langle - \widehat{n} \times \overline{I}_b \left(\widehat{t}_i^b f_n^b(\overline{r}')\right) \, , \, \widehat{t}_\ell^\rho \omega_m^\rho(\overline{r}) \right\rangle_{S_\rho} \quad ; \quad \rho = c, d \tag{44c}$$

where

$$n = r + (i-1)M_b \quad ; \quad 1 \le r \le M_b \quad ; \quad 1 \le i \le 2 \tag{44d}$$

and

$$m = q + (\ell - 1)M_\rho \quad ; \quad 1 \le q \le M_\rho \quad ; \quad 1 \le \ell \le 2 \tag{44e}$$

Finally,

$$Z_{nm}^{cd} = \mu_0 \left\langle \hat{n} \times \overline{I}_d \left(\hat{t}_\ell^d f_m^d (\overline{r}') \right) \, , \, \hat{t}_i^c \omega_n^c (\overline{r}) \right\rangle_{S_c} \tag{45a}$$

$$Z_{mn}^{dc} = -\left\langle \hat{n} \times \overline{I}_c \left(\hat{t}_i^c f_n^c (\overline{r}') \right) \, , \, \hat{t}_\ell^d \omega_m^d (\overline{r}) \right\rangle_{S_d} \tag{45b}$$

where m is given in (44e) with $\rho = d$ and n is given in (44d) with M_b replaced by M_c.

ACKNOWLEDGEMENT

This work was supported in part by the Joint Services Electronics Program (contract N00014-89-j-1007) and by the Ohio State University Research Foundation.

The Guest Editor thanks D. L. Jaggard and one anonymous reviewer for reviewing the paper.

REFERENCES

1. Engheta, N., and D. L. Jaggard, "Electromagnetic chirality and its applications," *IEEE Antennas and Propagation Society Newsletter*, Vol. 30, 6–12, Oct. 1988.
2. Varadan, V. K., V. V. Varadan, and A. Lakhtakia, "On the possibility of designing anti-reflection coating using chiral composites," *J. Wave-Material Interaction*, Vol. 2, 71–81, Jan. 1987.
3. Eftimiu, C., and L. W. Pearson, "Guided electromagnetic waves in chiral media," *Radio Science*, Vol. 24, 351–359, May-June 1989.
4. Fedorov, F. I., "Theory of gyrotropy" (in Russian), *Nauka i Technika*, Minsk, 1976.
5. Lakhtakia, A., V. K. Varadan, and V. V. Varadan, *Time-Harmonic Electromagnetic Fields in Chiral Media*, Springer, Berlin, 1989.
6. Richmond, J. H., "Scattering by a dielectric cylinder of arbitrary cross-section shape," *IEEE Trans. Antennas Propagat.*, Vol. AP-13, 334–341, May 1965.
7. Kluskens, M. S., and E. H. Newman, "Scattering by a chiral cylinder of arbitrary cross section," accepted for publication in the *IEEE Trans. Antennas Propagat.*
8. Morita, N., "Surface integral representations for electromagnetic scattering from dielectric cylinders," *IEEE Trans. Antennas Propagat.*, Vol. AP-26, 261–266, March 1978.
9. Rojas, R. G., "Integral equations for the scattering by two and three dimensional inhomogeneous chiral bodies," presented at the 1990 URSI Radio Science Meeting, Dallas, Texas.
10. Bohren, C. F., "Scattering of electromagnetic waves by an optically active cylinder," *J. Colloid Interface Sci.*, Vol. 66, 105–109, Aug. 1978.
11. Bohren, C. F., "Scattering of electromagnetic waves by an optically active spherical shell," *J. Chem. Phys.*, Vol. 62, 1566–1571, Feb. 1975.
12. Bohren, C. F., "Light scattering by an optically active sphere," *Chem. Phys. Lett.*, Vol. 29, 458–462, Dec. 1974
13. Jaggard, D. L., A. R. Michelson, and C. H. Papas, "On electromagnetic waves in chiral media," *Appl. Phys.*, Vol. 18, 211–216, 1979.

14. Engheta, N., and P. Pelet, "Modes in chirowaveguides," *Optics Letters,* Vol. 14, 593–595, June 1989.
15. Engheta, N., and S. Bassiri, "One- and two-dimensional dyadic Green's functions in chiral media," *IEEE Trans. Antennas Propagat.,* Vol. 37, 512–515, April 1989.
16. Weiglhofer, W. S., "Isotropic chiral media and scalar Hertz potentials," *J. Phys. A,* Vol. 21, 2249–2251, 1988.
17. Bassiri, S., C. H. Papas, and N. Engheta, "Electromagnetic wave propagation through a dielectric-chiral interface and through a chiral slab," *J. Opt. Soc. Am. A,* Vol. 5, 1450–1459, Sept. 1988.
18. Jaggard, D. L., X. Sun, and N. Engheta, "Canonical sources and duality in chiral media," *IEEE Trans. Antennas Propagat.,* Vol. AP-36, 1007–1013, July 1988.
19. Lakhtakia, A., V. V. Varadan, and V. K. Varadan, "Field equations, Huygens's principle, integral equations, and theorems for radiation and scattering of electromagnetic waves in isotropic chiral media," *J. Opt. Soc. Am. A,* Vol. 5, 175–184, Feb. 1988.
20. Lakhtakia, A., V. V. Varadan, and V. K. Varadan, "Radiation by a straight thin-wire antenna embedded in an isotropic chiral media," *IEEE Trans. Electromagn. Compat.,* Vol. EMC-30, 84–87, Feb. 1988.
21. Lakhtakia, A., V. K. Varadan, and V. V. Varadan, "Scattering and absorption characteristics on lossy dielectric, chiral, nonspherical objects," *Applied Optics,* Vol. 24, 4146–4154, Dec. 1985.
22. Bassiri, S., N. Engheta, and C. H. Papas, "Dyadic Green's function and dipole radiation in chiral media," *Alta Frequenza,* Vol. LV, 83–88, Mar.–Apr. 1986.
23. Varadan, V. V., A. Lakhtakia, and V. K. Varadan, "Equivalent dipole moments of helical arrangements of small, isotropic, point-polarizable scatters: application to chiral polymer design," *J. Appl. Phys.,* Vol. 63, 280–284, Jan. 1988.
24. Tai, C. T., "A note on the integral equations for the scattering of a plane wave by an electromagnetically permeable body," *Electromagnetics,* Vol. 5, 79–88,1985.
25. Harrington, R. F., *Field Computation by Moment Method,* McMillan, New York, 1968.
26. Lindell, I. V., and A. H. Sihvola, "Quasi-static analysis of scattering from a chiral sphere," *Report 54,* Helsinki University of Technology, Faculty of Electrical Engineering, Nov. 1989.
27. Harrington, R. F., *Time Harmonic Electromagnetic Fields,* McGraw-Hill, New York, 1961.

Roberto G. Rojas received the B.S.E.E. degree from New Mexico State University in 1979, and the M.Sc. and Ph.D. degrees in electrical engineering from the Ohio State University in 1981 and 1985, respectively. Since 1979 he has been with the Ohio State University ElectroScience Laboratory, initially as a Graduate Research Associate, later as a Post-Doctoral Researcher, and currently holding the position of Senior Research Associate. His current research interests are in electromagnetic scattering and modeling, high frequency techniques, and microstrip antennas. Dr. Rojas has won the 1988 R.W.P. King Prize Paper Award, the 1990 Browder J. Thompson Memorial Prize Award, both given by IEEE, and the 1989 Research Award, given by the College of Engineering at the Ohio State University. Dr. Rojas has served as Chairman, Vice-Chairman, and Secretary/Treasurer of the Columbus, Ohio Chapter of the IEEE Antennas and Propagation and Microwave Theory and Techniques Societies. He is a Senior member of IEEE, a member of URSI Commission B, and was a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.

Bandwidth Performance of Linear Adaptive Arrays with Tapped Delay-Line Processing

The nulling bandwidth of adaptive arrays with tapped delay-line processing is examined. Linear arrays with up to 10 elements are considered. It is shown how the number of taps in the delay lines and the amount of delay between taps affect the nulling bandwidth. For each size of array, the optimal number of delay-line taps and the optimal intertap delays are determined as functions of the required nulling bandwidth.

Manuscript received November 7, 1990; revised September 19, 1991.

IEEE Log No. 9107204.

This work was supported in part by the Joint Services Electronics Program under Contract N00014-89-J-1007.

I. INTRODUCTION

The ability of an adaptive array to null interference depends strongly on the interference bandwidth [1]. In general, nulling performance drops quickly as interference bandwidth increases. A well-known technique for improving the bandwidth performance of adaptive arrays is to use tapped delay-lines behind the elements.

The use of tapped delay-lines in adaptive arrays was first suggested by Widrow, et al. [2], and has been examined in several subsequent papers. In an early study, Rodgers and Compton [3] evaluated the bandwidth performance of a two-element tapped delay-line array with real weights. Mayhan, Simmons, and Cummings [4] used an analysis based on the eigenvalues of the covariance matrix to discuss how the number of elements and the number of delay-line taps affect the bandwidth performance of an adaptive array. White [5] addressed the tradeoff between using tapped delay lines versus additional auxiliary elements in adaptive sidelobe cancellers. Finally, Compton [6] has described how the number of taps and the delay between taps affect the bandwidth performance for a simple two-element array when complex weights are used.

The purpose here is to extend the results in [6] to the case of a linear array with up to ten elements and an arbitrary (possibly unequal) number of taps behind each element. For such arrays we show below how the number of delay-line taps and the amount of delay between taps affect the nulling bandwidth. For each array size, we determine the optimal number of delay-line taps and the optimal intertap delay as functions of the required nulling bandwidth. The results in this paper are taken from Vook [7], and more detailed information may be found there.

In a previous paper [8] by one of the authors, it was shown that the nulling bandwidth of an adaptive array with fast Fourier transform (FFT) processing is identical to that of an array with the equivalent tapped delay-line processing. The equivalent tapped delay-line processor has the same number of taps in the delay lines as the number of samples used in the FFTs and has the same delay between taps as the sampling time in the FFTs. Because of this equivalence, the results here can also be used to find the optimal sampling time and optimal number of samples for arrays with FFT processing.

In Section II, we present the equations needed to calculate signal-to-interference-plus-noise ratio (SINR) for an M-element linear adaptive array with an arbitrary configuration of tapped delay lines behind the elements. In Section III we present results based on these equations and discuss the bandwidth performance of linear arrays with up to ten elements as a function of the number of taps and the tap

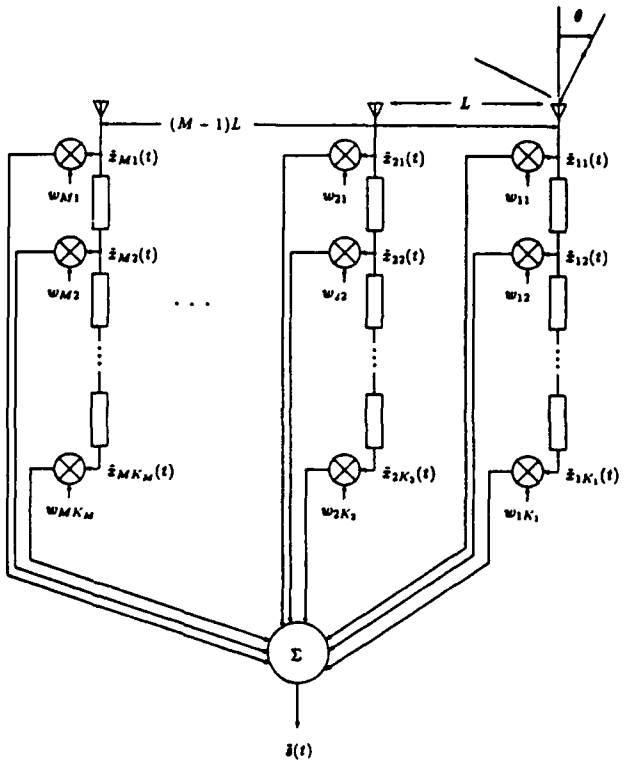


Fig. 1. M-element adaptive array with tapped delay-line processing.

spacing behind each element. Section IV contains our conclusions.

II. DEFINITIONS

Fig. 1 shows an M element adaptive array with tapped delay-line processing. The antenna elements are assumed isotropic and separated by a distance of one-half wavelength at the carrier frequency ω_0. Behind element m is a tapped delay line with $K_m - 1$ delays, K_m taps, and a time delay of T_0 seconds between taps. Different elements may have different numbers of taps in their delay lines. The *tap configuration* of the array is denoted by the vector $\mathbf{K}$, whose components are the number of taps behind elements 1 through M. That is $\mathbf{K} = (K_1, K_2, \ldots, K_M)$.

We assume that the array receives a desired signal and an arbitrary number of interfering signals and that each antenna element contains an independent thermal noise signal due to a preamplifier or mixer behind the element. The desired signal arrives from angle θ_d relative to broadside (θ is defined in Fig. 1). The interfering signals arrive from angles $\theta_{i_1} \ldots \theta_{i_N}$ where N is the number of interfering signals incident on the array.

It is assumed that the desired signal $\tilde{d}(t)$ is a zero-mean, stationary, random process of average power $p_d = E[|\tilde{d}(t)|^2]$, and that $\tilde{d}(t)$ has a flat, bandlimited power spectral density $S_{\tilde{d}}(\omega)$ equal to $2\pi p_d/\Delta\omega_d$ over a bandwidth $\Delta\omega_d$ centered at the carrier frequency ω_0. We denote the fractional bandwidth of the desired signal by $B_d = \Delta\omega_d/\omega_0$.

The N interfering signals are assumed to be mutually independent, zero-mean, stationary, random processes independent of $\tilde{d}(t)$ with average power $p_{i_l} = E[|\tilde{i}_l(t)|^2]$. Like the desired signal, the lth interfering signal is assumed to have a flat, bandlimited power spectral density $S_{\tilde{i}_l}(\omega)$ equal to $2\pi p_{i_l}/\Delta\omega_{i_l}$ over a bandwidth $\Delta\omega_{i_l}$, centered at frequency ω_0. The fractional bandwidth of the lth interfering signal is denoted by $B_{i_l} = \Delta\omega_{i_l}/\omega_0$.

Finally, each element signal is assumed to contain a zero mean thermal noise signal $\tilde{n}_{m1}(t)$ that is independent of the desired and interfering signals. The noise signals on different elements are assumed to be independent and to have average power σ^2. The noise power spectral density $S_{\tilde{n}}(\omega)$ is assumed equal to $2\pi\sigma^2/\Delta\omega_n$ over a bandwidth $\Delta\omega_n$, centered at ω_0. The fractional bandwidth of the noise signal is then given by $B_n = \Delta\omega_n/\omega_0$.

Let $\tilde{x}_{mk}(t)$ be the received signal at tap k behind element m. Each tap signal $\tilde{x}_{mk}(t)$ is multiplied by a complex weight w_{mk} and then summed to form the array output. The weight vector $\mathbf{W} = [w_{11}, w_{12}, \ldots, w_1 = w_{1K_1}, \ldots, w_{MK_M}]^T$ giving maximum output SINR for a given set of tap signals can be calculated using the method described in [6]. The optimal weight vector is given by [1]

$$\mathbf{W} = \Phi^{-1}\mathbf{S} \tag{1}$$

where Φ is the covariance matrix

$$\Phi = E[\mathbf{X}^*\mathbf{X}^T] \tag{2}$$

and $\mathbf{S}$ is the steering vector,

$$\mathbf{S} = E[\mathbf{X}^*\tilde{d}_0(t)]. \tag{3}$$

In these equations, $\mathbf{X} = [\tilde{x}_{11}(t), \ldots, \tilde{x}_{MK_M}(t)]^T$ is the signal vector and $\tilde{d}_0(t)$ is a replica of the desired signal waveform. The covariance matrix Φ and the steering vector $\mathbf{S}$ are computed as described in [6] except with more array elements and possibly an unequal number of taps per element. Once $\mathbf{W}$ has been computed for a given signal scenario, the output SINR is given by

$$\text{SINR} = \frac{P_d}{P_n + P_{(i_1)} + P_{(i_2)} + \cdots + P_{(i_N)}}. \tag{4}$$

P_d, P_n, and $P_{(i_l)}$ are, respectively, the output desired, noise, and interfering powers [6].

Several normalized parameters are used in the next section. As described in [6], the time delay between taps, T_0, may be expressed in normalized form

$$T_0 = rT_{90} = \frac{r\pi}{2\omega_0} \tag{5}$$

where T_{90} is the time delay required to produce a 90° phase shift at ω_0 (i.e., T_{90} is a quarter-wave delay at ω_0), and r is the number of quarter wave delays in T_0.

The signal-to-noise ratios (SNRs) are denoted by,

$$\xi_d = \frac{p_d}{\sigma^2} = \text{desired SNR per element} \qquad (6)$$

and

$$\xi_{i_l} = \frac{p_{i_l}}{\sigma^2}$$

$$= \text{INR per element for interfering signal } l,$$

$$l = 1,\ldots,N. \qquad (7)$$

III. BANDWIDTH PERFORMANCE OF ADAPTIVE ARRAYS

We now consider how tapped delay-line processing affects the bandwidth performance of the adaptive array. For convenience, we assume all signals received by the array have the same fractional bandwidth $B = B_d = B_{i_l} = B_n$. We also assume the desired signal arrives from broadside ($\theta_d = 0°$) with an SNR per element of 0 dB.

In the figures below, we plot the output SINR of each array as a function of the arrival angle of one interfering signal with the arrival angles of the other interfering signals held fixed. In general, the choice of interference angles often has a large effect on the performance. To evaluate the bandwidth performance, we compare the array output SINR with the SINR obtained with zero bandwidth (CW) signals. For easy comparison, the SINR for CW signals is included in every figure.

An M element array has $M-1$ degrees of freedom, so the number of pattern nulls that can be arbitrarily steered is at most $M-1$ [1]. In this study, for an M element array, we assume there are always $M-1$ interfering signals so that the array is in its most stressed condition. In general, the stronger the power of an interfering signal, the more sensitive the array is to the interference bandwidth [1]. In the results presented below, we assume all interfering signals have the same INR per element, 40 dB.

A. Number of Taps

We begin by showing a few examples. Fig. 2 shows the SINR for a ten-element array with no tapped delay-line processing. In this figure, there are eight fixed interfering signals located every 20° from −70° to +70°. The SINR is plotted as a function of θ_{i_1}, the ninth interfering signal angle, for increasing values of B. The larger the bandwidth, the worse the SINR. For $B = 0.1$, the degradation is as large as 25 dB for some values of θ_{i_1}. The array performance under these conditions is clearly unacceptable. Note the sharp spikes in the SINR curves, which result when two interfering signals arrive from the same angle so that only one pattern null is required to null both signals.

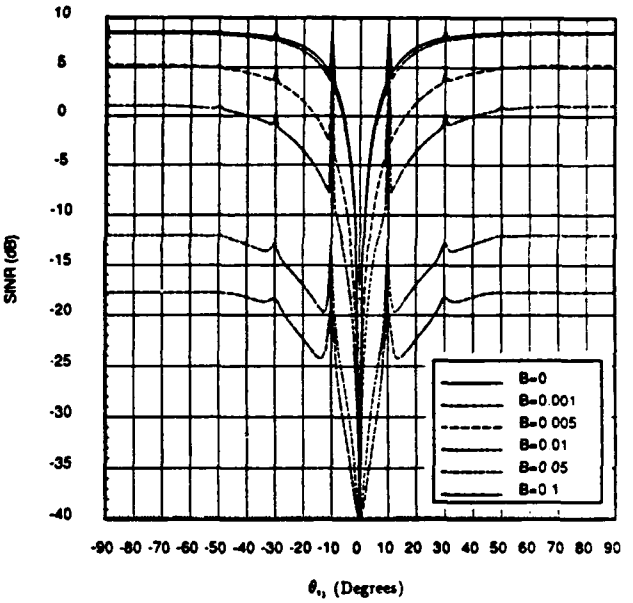


Fig. 2. SINR versus θ_{i_1}, 10-element array, 1 tap per element.

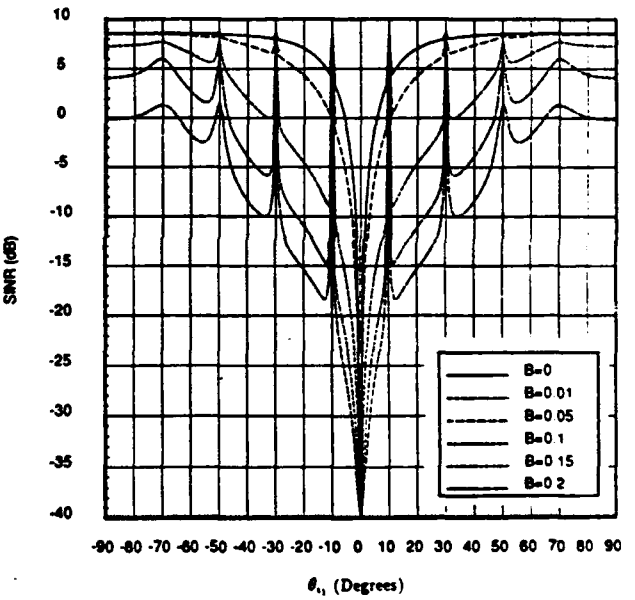


Fig. 3. SINR versus θ_{i_1}, 10-element array, 2 taps per element.

Next, Fig. 3 shows the SINR when a single quarter wave delay and an extra weight are added behind every element, with all other parameters unchanged from Fig. 2. With two taps per element, the SINR is equivalent to the CW SINR for bandwidths up to 0.01. Fig. 4 shows the SINR if another quarter wave delay and another weight is added to every element. *Note that with three taps per element, the SINR is within 1 dB of that obtained with CW interference for bandwidths up to 0.1.*

In the preceding examples, the SINR was plotted for increasing values of B while the number of taps per element was fixed. Now we show an example where the bandwidth is fixed and the number of taps

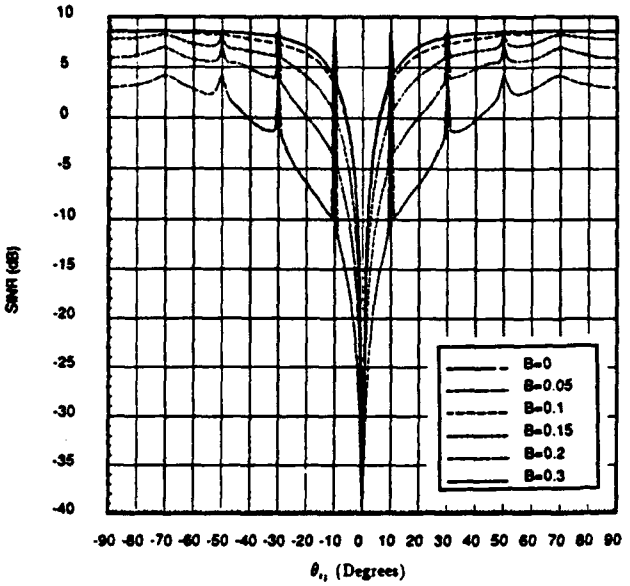


Fig. 4. SINR versus θ_{i_1}, 10-element array, 3 taps per element.

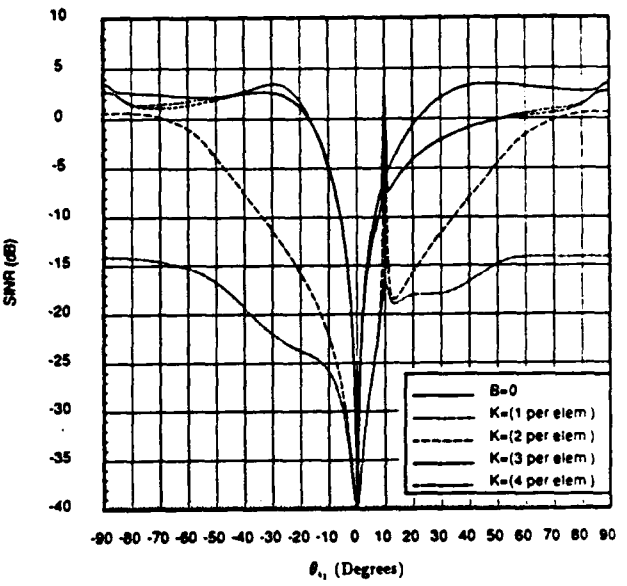


Fig. 6. SINR versus θ_{i_1}, 5-element array, $r = 1$.

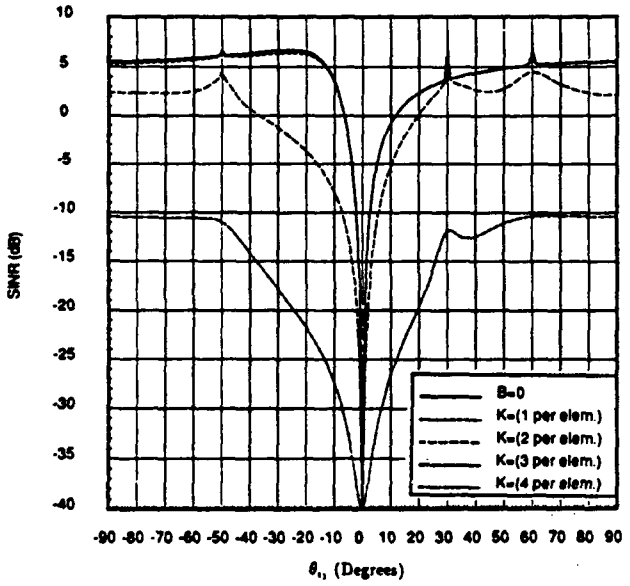


Fig. 5. SINR versus θ_{i_1}, 5-element array, $r = 1$.

per element is varied. Fig. 5 shows the SINR of a five-element array that receives signals of bandwidth 0.2. The fixed interfering signals arrive from $-10°$, $+30°$, and $+60°$ and the SINR is plotted for 1, 2, 3, and 4 taps per element. Note that the array needs at least three taps per element to recover the CW performance for all values of θ_{i_1} in the figure.

Now consider what happens when the fixed interfering signals for Fig. 5 are changed to $-90°$, $10°$, and $90°$. This case is plotted in Fig. 6. We again find that the maximum SINR is obtained with three taps per element, but there is a difference from the last case. In Fig. 6, for some θ_{i_1} the SINR for $B = 0.2$ is never as high as with CW interference, regardless of the number of taps per element. (For θ_{i_1} between $10°$ and $80°$, the maximum attainable SINR with any number of taps per element is between 2 and 4 dB below the CW curve.) This example illustrates a case in which it is not possible to recover the CW performance by using tapped delay-lines. However, in both Figs. 5 and 6, the maximum *attainable* SINR for $B = 0.2$ is achieved with three taps per element.

For a three-element array (with two interfering signals), it is always possible to recover the CW performance by using a sufficient number of taps per element, regardless of the interference angles. For arrays with more than three elements, however, one can find examples where the maximum attainable SINR with tapped delay-lines is less than the SINR for CW signals.[1] In these cases, as more taps are added, the SINR reaches a plateau as much as 5 dB below the CW SINR. These cases typically occur for large fractional bandwidths ($B > 0.1$) and when one or two interfering signals arrive from $+90°$ or $-90°$. We have found no cases in which the attainable SINR was more than 5 dB below the CW SINR.

It may appear that this effect is caused by too many interfering signals incident on the array. With $M - 1$ interfering signals, the array does not have enough degrees of freedom to place a pattern maximum on the desired signal. To test this possibility, we show in Fig. 7 the SINR for the same array as in Fig. 6 when we remove the fixed interfering signal at $10°$, so a total of $M - 2$ interfering signals remain. In this case, we find that as we add taps, the maximum performance still does not reach the CW performance

[1]There are also a few cases where the SINR with non-zero bandwidth interference is actually *higher* than with CW interference. We do not explore this behavior here.

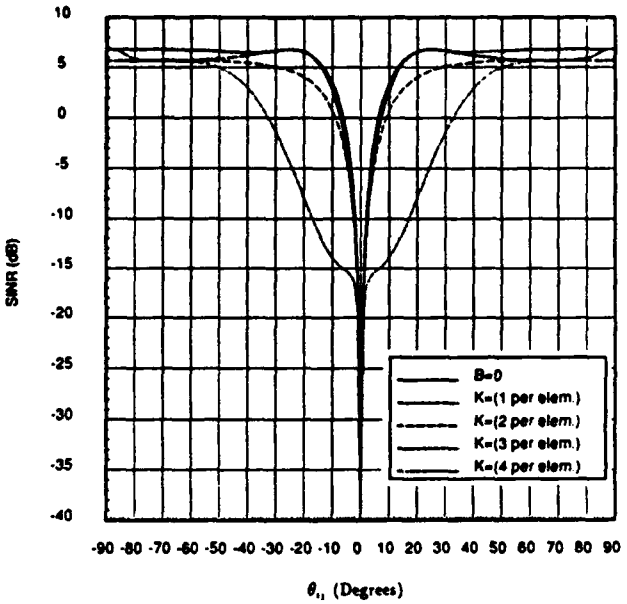


Fig. 7. SINR versus θ_{i_1}, t-element array, $r = 1$.

for θ_{i_1} between 40° and 80°. The maximum attainable SINR is approximately 1.1 dB less than the CW SINR for θ_{i_1} in that range. Thus, reducing the number of interfering signals can still result in a situation where the maximum SINR is less than the CW SINR.

We define the optimal number of taps per element for a given bandwidth to be the minimum number of taps required to be within 1 dB of the attainable SINR. (In this definition, we also assume $r = 1$. Recall that r is the delay between taps in quarter wavelengths defined in (5). As is discussed below, the value $r = 1$ is always within the range where maximum SINR is obtained.)

We have examined numerous SINR curves of the type shown in Figs. 2–7. A summary of the results is shown in Fig. 8, which plots the optimal number of taps per element versus the signal bandwidth for 3 to 10 elements.[2] For example, a seven-element array receiving signals with $B = 0.1$ should, for best performance, have three taps per element. In general, as the number of elements increases, the optimal number of taps per element increases for a given bandwidth.

As seen in Figs. 5 and 6, interference angles have a large effect on the performance. Some combinations of θ_{i_l}, $l = 1, \ldots, M-1$, yield poorer bandwidth performance than others. The optimal taps per element plotted in Fig. 8 take into account the arrival angles that cause the largest SINR degradation.

[2]For arrays with FFT processing, Fig. 8 shows the optimal (i.e., minimum) number of samples required in the FFT processors. Since it is easiest to use 2^n samples in an FFT processor, where n is some integer, one would use the smallest n such that 2^n is at least as large as the value shown in Fig. 8.

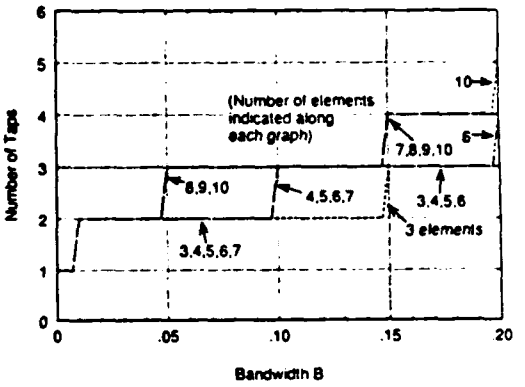


Fig. 8. Required number of taps per element versus bandwidth B.

B. Unequal Number of Taps per Element

The preceding discussion dealt with arrays in which each element in the array has the same number of delay-line taps. It is also interesting to consider whether one could use a different number of taps behind different array elements to reduce the total number of weights needed in the array. To examine this question, we start with a three-element array to illustrate typical SINR curves and then summarize results for arrays with more than three elements.

For a given tap configuration in a three-element array, there is a maximum bandwidth at which the array can achieve the CW performance. We define the *bandwidth cutoff* B_c of an array to be the maximum bandwidth such that the array SINR is within 1 dB of its CW value for any $B \leq B_c$. (In this definition, we again assume that $r = 1$.) It is shown in [7] that for a three-element array with two taps per element, the bandwidth cutoff B_c is 0.1, but for three taps per element B_c is 0.5. Since there is such a large difference between these two values of B_c, the question arises whether one could use fewer than three taps per element when the signal bandwidth is between 0.1 and 0.5. If $B = 0.2$, for example, must one add an extra tap to *every* element of the array?

Table I gives the bandwidth cutoff for a three-element array with various tap configurations. For example, if the tap configuration is $\mathbf{K} = (2,3,2)$, then the array achieves the CW performance if the bandwidth is less than 0.35 for all combinations of two interfering signals. These values of B_c were determined by examining numerous SINR curves to see the full effects of different arrival angles. (Since our evaluation of bandwidth performance takes into account *all* combinations of interference angles, symmetrical tap configurations such as $\mathbf{K} = (2,2,3)$ and $\mathbf{K} = (3,2,2)$ have the same bandwidth cutoff.)

An interesting result shown in Table I is that the bandwidth cutoff for $\mathbf{K} = (2,1,2)$ is the same as that for $\mathbf{K} = (2,1,1)$, but for $\mathbf{K} = (2,2,1)$ B_c is about ten times higher. For an array with $\mathbf{K} = (2,1,2)$ and bandwidths in the range of 0.005 to 0.05, it turns out

TABLE I
Bandwidth Cutoffs for Three-Element Array

	Bandwidth Cutoff			
	K	B_c	K	B_c
3-Element Array	(1,2,1)	0.005	(2,3,1)	0.07
2 Interfering Signals	(2,1,1)	0.005	(3,1,2)	0.005
$\xi_{i_1} = \xi_{i_2} = 40$ dB	(2,1,2)	0.005	(3,2,2)	0.2
$\xi_d = 0$ dB	(2,2,1)	0.05	(2,3,2)	0.35
$r = 1$	(2,2,2)	0.1	(3,3,3)	0.5

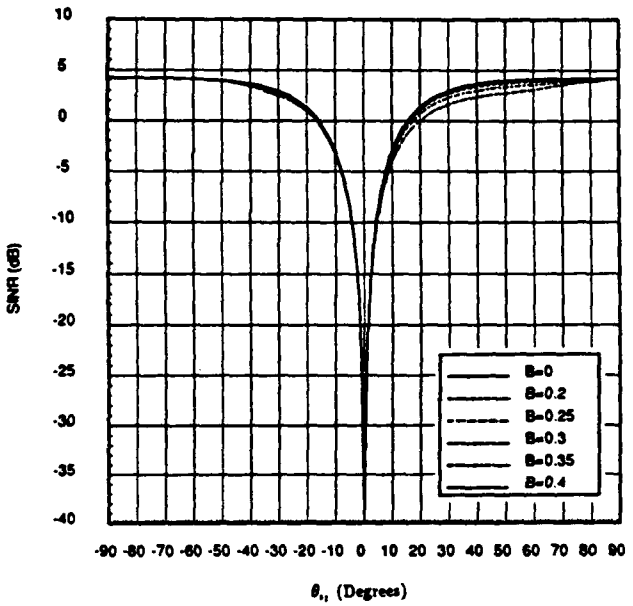


Fig. 9. SINR versus θ_{i_1}, 3-element array, K = (2,3,2), $r = 1$.

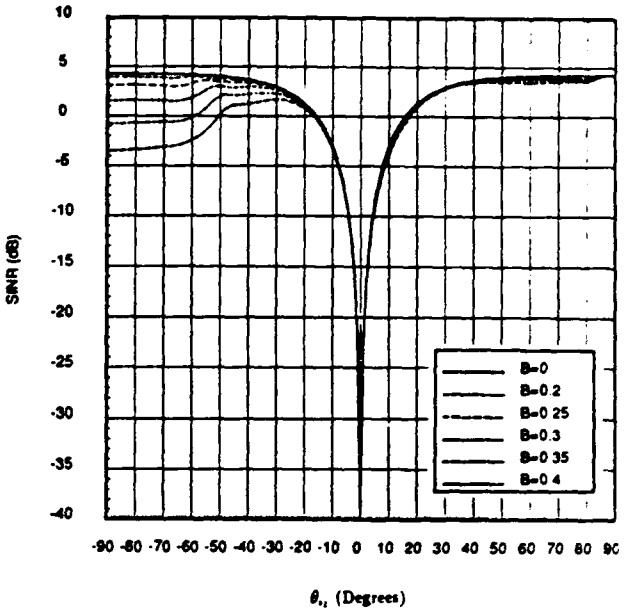


Fig. 10. SINR versus θ_{i_1}, 3-element array, K = (3,2,2), $r = 1$.

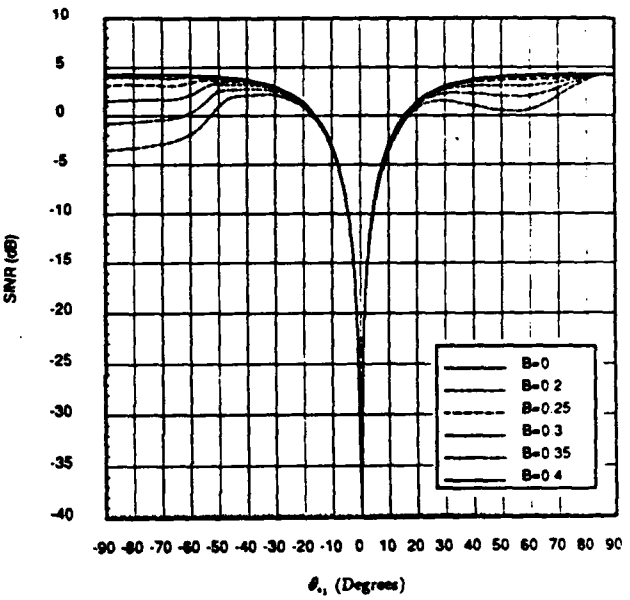


Fig. 11. SINR versus θ_{i_1}, 3-element array, K = (2,2,3), $r = 1$.

that certain combinations of interference arrival angles can cause large (> 10 dB) drops in SINR. However, an array with K = (2,2,1) does not have these drops in SINR and is still capable of optimal performance in that range of bandwidth. Since rearranging the taps from K = (2,1,2) to K = (2,2,1) allows the array to handle much higher bandwidths, this result suggests that an extra tap should be placed behind a middle element instead of an outer element for best performance.

Also note that there is no performance advantage in choosing tap configurations such as K = (3,1,2) or K = (2,3,1) over K = (2,2,2). Although these three configurations have the same total number of taps (i.e., 6), K = (2,2,2) yields a bandwidth cutoff that is clearly higher than the other two. This result suggests that it is better to divide the taps equally among the elements than to have one element with two more taps than another element.

In Figs. 9, 10, and 11, we plot the SINR of three-element arrays with K = (2,3,2), K = (3,2,2), and K = (2,2,3), respectively, for different bandwidths. The fixed interfering signal in these three figures arrives from 90°. With a tap configuration of either K = (2,2,3) or K = (3,2,2) the array achieves optimal performance for signal bandwidths up to 0.2. However, a tap configuration of K = (2,3,2) yields maximum performance for bandwidths up to 0.35. An extra tap on the middle element again yields better results than an extra tap on the end.

For arrays with more than three elements, the results are similar. For example, one finds that a four-element array with K = (2,3,3,2) has a B_c of 0.25, but when K = (3,2,2,3), the B_c is 0.15. (With more than three elements, we use maximum attainable SINR rather than CW SINR to define B_c.)

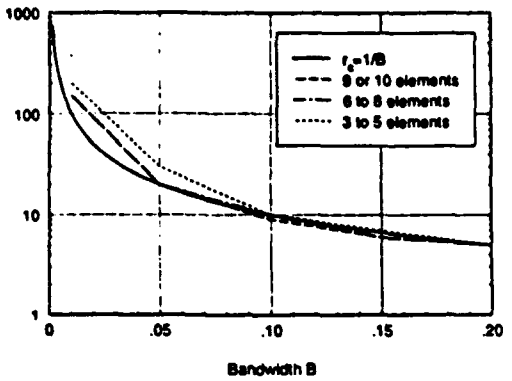


Fig. 12. Intertap delay cutoff r_c versus bandwidth B.

The following line of reasoning appears to explain why it is preferable to add an extra tap to a middle element. An arbitrary signal received from an angle $\theta > 0$ will reach element 1 first, element 2 after some time delay T, and finally element 3 after a delay $2T$. The tapped delay-lines compensate for the decorrelation due to T by delaying the signals on one element so that they are more correlated with signals on other elements. An extra tap on an outer element can do this only when the outer element receives the signal first. When $\theta > 0$, a signal on element 3 is a delayed version of the signal on elements 1 and 2. An extra tap on element 3 will not restore the correlation between the element 3 signals and the signals on the other elements. Placing the extra tap on an outer element improves the performance mainly when the interfering signals reach that element before the others. However, placing the extra tap on the middle element improves the correlation of the middle element signal with the signal on at least one other element for any θ. Recall that our evaluation of bandwidth performance takes into account *all* arrival angles of the interference. For this reason, an array with the extra tap on an outer element does not perform as well as one with the extra tap on the middle element when all interfering angles are taken into account.

Note from Figs. 10 and 11 that the performance of arrays with $\mathbf{K} = (3,2,2)$ or $\mathbf{K} = (2,2,3)$ drops primarily when one or both of the interfering signals are being received first by the two-tap outer element (see Fig. 1). (Recall that $\theta_{i_2} = 90°$ in both Figs. 10 and 11.) In Fig. 10, when $\theta_{i_1} > 0°$, both interfering signals are received first by the three-tap outer element, and there is very little drop in SINR for $B < 0.4$. However, in Fig. 11, when $\theta_{i_1} > 0°$ both interfering signals are received first by the two tap outer element, and there is a significant drop in SINR for $B > 0.3$. When $\theta_{i_1} = -90°$ in Figs. 10 and 11, one of the two interfering signals is received first by the two-tap outer element, and the SINR drops accordingly. Fig. 9 shows that the SINR of an array with $\mathbf{K} = (2,3,2)$ does not have these drops in SINR for the same bandwidth range.

Thus, the results in Table I suggest that for best SINR performance, the taps in the array should be shared as equally as possible among the elements. Any leftover taps should be placed on the elements near the physical center of the array.

C. Delay Between Taps

Throughout this discussion, we have assumed a quarter wavelength delay ($r = 1$) between taps at the carrier frequency. Let us now consider how the SINR is affected by the delay between taps. We limit the discussion here to the case where the array has the optimal number of taps per element.

In [6], it was shown that the SINR performance of a two-element array is rather insensitive to the value of r as long as r is kept below the value of $1/B$. For two elements and two taps per element, any value of r less than $1/B$ yields optimal performance. Values of r between $1/B$ and $4/B$ result in suboptimal performance with two taps per element, but optimal performance can be recovered if more taps are added to both elements. Values of r greater than $4/B$ result in suboptimal performance regardless of how many extra taps are added.

We find similar behavior for arrays with up to ten elements. For larger arrays, our results indicate that the SINR with an optimal number of taps is maximum and independent of r as long as r is less than a certain maximum value. Above this maximum value, array performance drops. This behavior is similar to the two-element case except that the maximum value of r is not simply $1/B$ for all cases. We determine the maximum value of r as follows.

For a given number of array elements and a given bandwidth, we first determine the optimal number of taps per element, as described above, with $r = 1$. Then, using this number of taps, we increase r and calculate the change in output SINR. There is a broad range of r over which the SINR is essentially constant, but if r is increased enough, the SINR finally begins to drop. We define the *intertap delay cutoff* r_c to be the value of r at which the SINR has dropped 1 dB from its optimal value at $r = 1$. Examples showing how r affects performance can be found in [7].

For arrays with more than two elements, r_c still depends mainly on the bandwidth of the signals, but also to a lesser extent on the number of elements in the array. Fig. 12 plots the intertap delay cutoffs versus the signal bandwidth for arrays with up to 10 elements. The number of taps per element in each case is the optimal value shown in Fig. 8. These results were obtained by examining numerous SINR curves for many combinations of interference angles. The interference angles often affect the value of r_c, and the values of r_c plotted in Fig. 12 are those for the worst choice of arrival angles. As a rule-of-thumb, it may be seen from Fig. 12 that the maximum intertap delay r_c

is usually around $1/B$. With six to ten elements and higher signal bandwidths, the value of r_c sometimes is slightly less than $1/B$.

With $r > r_c$, array performance is usually similar to that described in [6]. For values of r up to approximately $4/B$ ($r_c < r < 4/B$), the performance of the array can be improved by adding more taps behind the elements. When r is greater than $4/B$, full performance cannot be achieved no matter how many taps are added. Examples of this behavior may be found in [7].

For $r < 1$, the behavior of the array is similar to that described in [6]. As r is reduced below 1, the SINR retains its optimal value but the weight magnitudes become large. As long as $r \neq 0$, the array can, with large weight magnitudes, achieve the necessary magnitude and phase response for interference rejection. However, small values of r result in an ill-conditioned covariance matrix, which may cause difficulty with the weight control algorithm.

IV. CONCLUSION

This paper has discussed how the nulling bandwidth of a linear adaptive array is affected by the tapped delay-line configuration behind the elements. Results were presented for arrays with up to ten elements.

To retain maximum SINR performance, an M element adaptive array receiving $M-1$ interfering signals generally requires more and more taps per element as the signal bandwidth increases. For best array performance with a fixed total number of taps, the elements in the array should share the taps as equally as possible. Any extra taps are best used on elements near the center of the array. The optimal number of taps per element for selected bandwidths can be found from Fig. 8.

Adaptive array performance is relatively insensitive to the delay between taps. If the array has the optimal number of taps per element for a given bandwidth, then any value of intertap delay less than the r_c shown in Fig. 12 (the values of r_c are in quarter wavelengths) will yield an SINR within 1 dB of the maximum attainable SINR. As a rule-of-thumb, the value of r_c is about $1/B$, so smaller bandwidths allow larger intertap delays. If intertap delays larger than r_c are used, the array output SINR drops. For intertap delays between $1/B$ and $4/B$, the performance can be restored by adding more taps behind the elements. For intertap delays larger than $4/B$, the array performance is suboptimal regardless of the number of taps for most interference angles.

F. W. VOOK
R. T. COMPTON, JR.
ElectroScience Laboratory
Dept. of Electrical Engineering
The Ohio State University
2015 Neil Ave.
Columbus, OH 43212

REFERENCES

[1] Compton, R. T., Jr. (1988)
Adaptive Antennas—Concepts and Performance.
Englewood Cliffs, NJ: Prentice-Hall, 1987.

[2] Widrow, B., Mantey, P. E., Griffiths, L. J., and Goode, B. B. (1967)
Adaptive antenna systems.
Proceedings of the IEEE, 55, 12 (Dec. 1967), 2143.

[3] Rodgers, W. E., and Compton, R. T., Jr. (1979)
Adaptive array bandwidth with tapped delay-line processing.
IEEE Transactions on Aerospace and Electronic Systems, AES-15, 1 (Jan. 1979), 21.

[4] Mayhan, J. T., Simmons, A. J., and Cummings, W. C. (1981)
Wideband adaptive antenna nulling using tapped delay-lines.
IEEE Transactions on Antennas and Propagation, AP-29, 6 (Nov. 1981), 923.

[5] White, W. D. (1983)
Wideband interference cancellation in adaptive sidelobe cancellers.
IEEE Transactions on Aerospace and Electronic Systems, AES-19, 6 (Nov. 1983), 915.

[6] Compton, R. T., Jr. (1988)
The bandwidth performance of a two-element adaptive array with tapped delay-line processing.
IEEE Transactions on Antennas and Propagation, 36, 1 (Jan. 1988), 5.

[7] Vook, F. W. (1989)
The bandwidth performance of adaptive arrays with tapped delay-line processing.
M.Sc. thesis, The Ohio State University, Dept. of Electrical Engineering, Columbus, OH, Mar. 1989.

[8] Compton, R. T., Jr. (1988)
The relationship between tapped delay-line and FFT processing in adaptive arrays.
IEEE Transactions on Antennas and Propagation, 36, 1 (Jan. 1988), 15.

[9] Applebaum, S. P. (1976)
Adaptive arrays.
IEEE Transactions on Antennas and Propagation, AP-24, 5 (Sept. 1976), 585.

Electromagnetic Field Excited by a Line Source Placed at the Edge of an Impedance Wedge

Giuseppe Pelosi, Roberto Tiberio, and Roberto G. Rojas

Abstract—A uniform high-frequency solution is presented for the field of a line source located at the edge of an impedance wedge. Surface wave contributions are explicitly included into the expression for the total field, which is continuous across their shadow boundaries.

I. Introduction

In several practical applications, a convenient location for wire and slot antennas is found near and at the tip of an edge joining two almost flat surfaces. In modern technology, such surfaces often consist of composite materials, which may be usefully modeled by surface impedance boundary conditions. In order to analyze these configurations, the canonical problem of an impedance wedge excited by a source located at its edge is relevant.

Recently, a high-frequency solution has been presented for the diffraction by a wedge with surface impedance faces, when the source and the observation points are located at finite distances from its edge [1]. There high-frequency expressions for the total field are given in the format of the uniform GTD (UTD). In this communication, the above formulation is extended to describe the scattering phenomenon when a line source is placed right at the edge of the impedance wedge. Also in this case, the solution to this canonical problem is obtained in a rigorous and very simple fashion, by employing a plane wave spectrum formulation. The same problem was treated earlier [2], [3]. However, both the approach and the formulation presented here are different. There, although the methodology is basically the same, the case of the tip excited wedge is handled as an object which is different from that of a plane wave excited wedge. Also, the solutions are rather involved. At a variance, here it is shown that the solution for the case we are presently concerned is directly and simply obtained from that for a plane wave illumination [4], when a spectrum representation of the field is used. Furthermore, our result is uniformly valid at any aspect of observation.

The conceptual generality of a plane wave spectrum representation suggests that the formulation presented here provides a basic step for further extension to treat the three-dimensional (3-D) case of a point source located at the edge of a wedge.

The organization of the presentation is summarized hereafter. First, an exact integral representation is obtained for the total field (Section III). Next, after extracting the surface wave, contributions the integral is asymptotically evaluated to give a uniform high-frequency solution (Section IV). The presence of complex poles is properly accounted for, so that the expression for the total field is continuous across the shadow boundaries of the surface waves [5]–[7]. Numerical results are presented in Section V.

II. Formulation

The geometry for the scattering at the edge of an impedance wedge is depicted in Fig. 1. A line source is located at Q and the

Manuscript received March 5, 1990; revised November 29, 1990. This work was supported in part by the Joint Service Program under Contract N00014-89-1007 and by The Ohio State University Research Foundation.

G. Pelosi and R. Tiberio are with the Department of Electronics Engineering, University of Florence, 50139 Florence, Italy.

R. G. Rojas is with the ElectroScience Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, OH 43212.

IEEE Log Number 9100273.

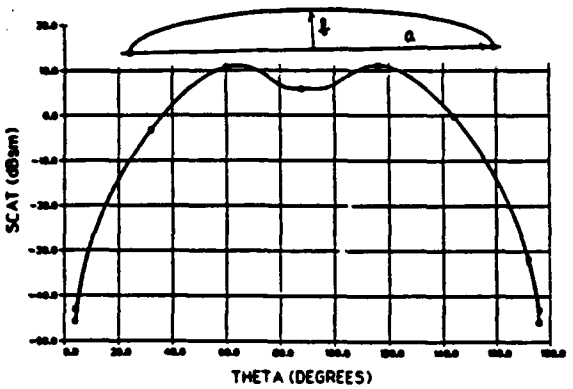


Fig. 2. Monostatic RCS for an anisotropic ellipsoid of revolution.

and that $\eta_{\theta\theta}$ and $\eta_{\phi\phi}$ are independent of the azimuthal angle. Then the backscattered field of the body S in the axial direction will be zero.

The proof of this theorem follows exactly that given in [1] and therefore will be omitted.

III. Numerical Demonstration of the Scattering Theorems

In the following, a numerical demonstration of the two theorems given above is presented. In producing these numerical calculations we modified a code for PEC scatterers written in the McDonnell Douglas Research Laboratories [2] to include an anisotropic surface impedance boundary condition. In Fig. 2 we show the monostatic H- and V-pol RCS for an ellipsoid of revolution. The two surface impedances ($\eta_{\theta\theta}$ and $\eta_{\phi\phi}$) are chosen to be imaginary. The condition of Theorems 1 and 2 are satisfied. The coincidence of the two RCS curves serve to demonstrate Theorem 1, while the zero RCS in the axial directions ($\theta = 0$ and 180°) serve to demonstrate Theorem 2.

IV. Conclusion

We have proved two RCS theorems in this communication for anisotropic surface boundary conditions. They constitute an extension of Weston's two theorems because we require only that $\eta_{\theta\theta}\eta_{\phi\phi} = 1$ instead of $\eta_{\theta\theta} = 1$ and $\eta_{\phi\phi} = 1$. We have also demonstrated, through some numerical examples, the truth of these theorems. We believe that the theorems can serve to check new MoM-BOR-IBC codes.

Acknowledgment

Our physicist colleagues, David Ingham and Gary Nizzi, tried to convince us, on physical ground, the truth of the generalized Weston theorems. This communication is the result of our attempt to understand in detail these facts which we observe from our code outputs. We appreciate helpful discussions with them. We thank Dr. John Prodan for useful discussions. Finally, we are very grateful that Ken Yamada shares with us and understands our working philosophy through which the writing of this communication is possible.

References

[1] V. H. Weston, "Theory of absorbers in scattering," *IEEE Trans. Antennas Propagat.*, vol. AP-11, pp. 578–584, Sept. 1963.

[2] J. M. Putnam and L. N. Medgyesi-Mitschang, "Combined field integral formulation for axially inhomogeneous bodies of revolution. Vol. I: Final report," McDonnell Douglas Res. Labs. Dec. 1987.

Electromagnetic Field Excited by a Line Source Placed at the Edge of an Impedance Wedge

Giuseppe Pelosi, Roberto Tiberio, and Roberto G. Rojas

Abstract—A uniform high-frequency solution is presented for the field of a line source located at the edge of an impedance wedge. Surface wave contributions are explicitly included into the expression for the total field, which is continuous across their shadow boundaries.

I. Introduction

In several practical applications, a convenient location for wire and slot antennas is found near and at the tip of an edge joining two almost flat surfaces. In modern technology, such surfaces often consist of composite materials, which may be usefully modeled by surface impedance boundary conditions. In order to analyze these configurations, the canonical problem of an impedance wedge excited by a source located at its edge is relevant.

Recently, a high-frequency solution has been presented for the diffraction by a wedge with surface impedance faces, when the source and the observation points are located at finite distances from its edge [1]. There high-frequency expressions for the total field are given in the format of the uniform GTD (UTD). In this communication, the above formulation is extended to describe the scattering phenomenon when a line source is placed right at the edge of the impedance wedge. Also in this case, the solution to this canonical problem is obtained in a rigorous and very simple fashion, by employing a plane wave spectrum formulation. The same problem was treated earlier [2], [3]. However, both the approach and the formulation presented here are different. There, although the methodology is basically the same, the case of the tip excited wedge is handled as an object which is different from that of a plane wave excited wedge. Also, the solutions are rather involved. At a variance, here it is shown that the solution for the case we are presently concerned is directly and simply obtained from that for a plane wave illumination [4], when a spectrum representation of the field is used. Furthermore, our result is uniformly valid at any aspect of observation.

The conceptual generality of a plane wave spectrum representation suggests that the formulation presented here provides a basic step for further extension to treat the three-dimensional (3-D) case of a point source located at the edge of a wedge.

The organization of the presentation is summarized hereafter. First, an exact integral representation is obtained for the total field (Section III). Next, after extracting the surface wave, contributions the integral is asymptotically evaluated to give a uniform high-frequency solution (Section IV). The presence of complex poles is properly accounted for, so that the expression for the total field is continuous across the shadow boundaries of the surface waves [5]–[7]. Numerical results are presented in Section V.

II. Formulation

The geometry for the scattering at the edge of an impedance wedge is depicted in Fig. 1. A line source is located at Q and the

Manuscript received March 5, 1990; revised November 29, 1990. This work was supported in part by the Joint Service Program under Contract N00014-89-1007 and by The Ohio State University Research Foundation.

G. Pelosi and R. Tiberio are with the Department of Electronics Engineering, University of Florence, 50139 Florence, Italy.

R. G. Rojas is with the ElectroScience Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, OH 43212.

IEEE Log Number 9100273.

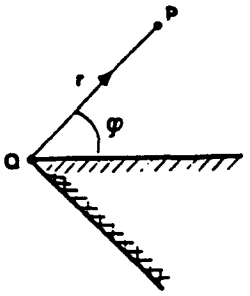


Fig. 1. Geometry for the scattering at the edge of an impedance wedge.

observation point at P; its location in a polar coordinate system with its origin at Q is denoted by (r, ϕ). The exterior angle of the wedge is $n\pi$. Two different uniform isotropic, arbitrary impedance boundary conditions may be imposed on the two faces. The surface impedances of the 0 and $n\pi$ faces are denoted by Z_0 and Z_n, respectively; $Z_{0,n}$ are complex numbers whose real part, because of energy considerations, must be non-negative. The following notation will be used to describe the boundary conditions for an electric (TM, e) and a magnetic (TE, h) line source

$$\sin\theta_p^e = Z_c/Z_p; \quad \sin\theta_p^h = Z_p/Z_c, \qquad p = 0, n \tag{1}$$

in which Z_c is the free-space impedance. Throughout the paper, superscripts e and h are suppressed; the TM and TE cases are treated together and the expressions presented later on apply to both cases, provided that the proper value for $\theta^{e,h}$ is used. A harmonic time dependence $\exp(j\omega t)$ is assumed and suppressed.

The procedure adopted in this communication may be summarized as follows.

1) First, by virtue of reciprocity a uniform line source is located at P and the observation point at the edge Q. The incident field from the source is represented by its plane wave spectrum.

2) Next, the plane wave response at Q of the wedge is found from the exact integral representation given in [4].

3) Then, an exact integral representation for the total field at Q is obtained by superposition of the above spectrum components. Thus again by reciprocity, it provides the desired expression for the field at P due to a line source at Q.

4) Finally, this integral representation is asymptotically evaluated to obtain our uniform high frequency solution.

Steps 1–3 are described in the next section and step 4 in Section IV.

III. Integral Representation

A basic step within the context of a plane wave spectrum representation, is that of obtaining the plane wave response $\tilde{u}$ of the wedge. Its expression for the field at Q is now found by using the exact solution given in [4]. A convenient expression for the total field $\tilde{u}$ at a point $Q' = (\rho, \alpha')$, due a unit plane wave incident on the edge of the wedge from a direction α, as depicted in Fig. 2, is

$$\tilde{u}(Q') = \frac{1}{2\pi j}\int_\gamma f(\xi)\, e^{jk\rho\cos\xi}\, d\xi \tag{2}$$

where γ is the Sommerfeld contour of integration, k is the propagation constant, and

$$f(\xi) = \frac{[S(\xi;\alpha,\alpha') - S(-\xi;\alpha,\alpha')]}{2} \tag{3}$$

is an odd function of ξ, in which $S(\xi;\alpha,\alpha')$ is the spectrum function obtained by Maliuzhinets [4], [1]. Expression (3) is simply a consequence of the symmetry of the contour γ; its usefulness is explained next.

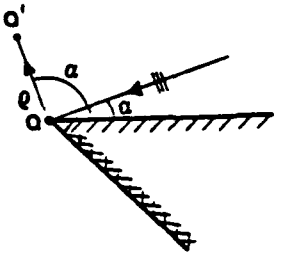


Fig. 2. Geometry for the plane wave response of the wedge.

It has been shown [8] that a transform relationship can be established between $F(\rho) = \tilde{u}(Q')$ and $f(\xi)$ in (2), via the transform pair

$$f(\xi) = \frac{\sin\xi}{2j}\int_0^\infty F(\rho)\, e^{-jk\rho\cos\xi}\, d\xi \tag{4}$$

and (2); provided that $f(\xi)$ is an odd function. Consequently, it is easily seen that

$$F(0) = \pm 2jf(\pm j\infty) \tag{5}$$

which yields the following exact expression for the total field at Q:

$$\tilde{u}(Q) = \frac{2}{n}g_n(\alpha) = \frac{1}{n}\frac{\mathscr{M}(\alpha)\sin(\alpha/n)}{\sin\left(\frac{\alpha+\theta_0}{2n}\right)\cos\left(\frac{\alpha-\theta_n}{2n}\right)} \tag{6}$$

where

$$\mathscr{M}(\alpha) = \frac{\Psi_n\left(n\pi-\alpha-\frac{\pi}{2}-\theta_0\right)\Psi_n\left(\alpha-\frac{\pi}{2}-\theta_n\right)}{\Psi_n\left(n\pi-\alpha-\frac{\pi}{2}+\theta_0\right)\Psi_n\left(\alpha-\frac{\pi}{2}+\theta_n\right)} \tag{7}$$

is a regular function, and Ψ_n is the Maliuzhinets special function [4]. In deriving (6), the asymptotic property of Ψ_n for $|\mathrm{Im}(\alpha)| \to \infty$ has been used. Also, it was found useful to employ the expression of the auxiliary function introduced by Maliuzhinets, which puts the complex electric poles into an explicit trigonometric form. Equation (6) is the desired plane wave response at Q of the wedge.

Let us now consider a unit, uniform line source located at P (Fig. 1) and the plane wave spectrum representation of its field. In order to determine its field at Q, first the above plane wave response (8) is analytically continued into complex space. Next, the spectrum of the field from the source at P is weighted by $\tilde{u}(Q)$, to give by superposition an exact integral representation for the total field u at the edge of the wedge

$$u = \frac{2e^{-j\pi/4}}{n\sqrt{2\pi}}\int_c g_n(\alpha)\, e^{-jkr\cos(\phi-\alpha)}\, d\alpha. \tag{8}$$

In (8) the contour of integration c is either $(-j\infty, \pi + j\infty)$ when $\phi \le n\pi/2$ or $(-\pi - j\infty, +j\infty)$ when $\phi \ge n\pi/2$.

By reciprocity, (8) also provides the required representation for the field u at any point P due to a line source at the edge Q. It is noted that when $\theta_0 = \theta_n$, u exhibits the expected symmetry for $P'(r, \phi)$ and $P'' = (r, n\pi - \phi)$. Furthermore, for a perfectly conducting wedge it is seen that

$$u = \frac{2}{n}\sqrt{\frac{\pi}{2}}\, e^{-j\pi/4}H_0^{(2)}(kr) \tag{9}$$

in the TE case ($\theta_0 = \theta_n = 0$) and $u = 0$ in the TM case. The field of an electric line source (TM) also vanishes when any one of the two faces is perfectly conducting. An uniform high-frequency approximation of u for kr large is presented in the next section.

Before proceeding further, it is worth pointing out that whenever the spectrum of a Sommerfeld integral representation of the response of a wedge is known for an arbitrarily incident plane wave (ϕ, β), the field $\tilde{u}(\phi, \beta)$ at a point Q on the edge is easily found by applying the concept of the transform pair (2) and (4). This can be done for example for impedance wedges with exterior angles $n\pi = 1, 1.5, 2\pi$ [9], and of course for a perfectly conducting wedge with any n. Then, this expression for $\tilde{u}(\phi, \beta)$ can be used to weight the double spectral integral representation of a source in the 3-D space. It provides the extension to the 3-D case of the formulation presented so far; however, this will be discussed in a subsequent paper and compared with [10].

IV. High-Frequency Solution

In order to obtain a high-frequency solution, it is useful to deform the contour of integration in (8) onto a steepest descent path (SDP) through the saddle point at $\alpha = \phi$. When a surface wave is excited on either faces of the wedge, a surface wave pole $\alpha = \eta_p - \phi$ $(p = 0, n; \eta_0 = -\theta_0, \eta_n = n\pi + \theta_n)$ may be captured in the contour deformation process if either $\phi < \alpha_0$ or $\phi > n\pi - \alpha_n$, in which

$$\alpha_p = -\theta_{pr} - gd(\theta_{pi})\,\mathrm{sgn}(\theta_{pi}) \tag{10}$$

where gd denotes the Gudermann function and $\theta_p = \theta_{pr} - j\theta_{pi}$. Thus, its residue contribution may occur in the solution, in addition to the contribution from the SDP integration.

Then, the integral along the SDP is asymptotically evaluated by applying the procedure suggested in [5]–[7], so that the crossing of a pole through the SDP is properly accounted for. It yields a uniform high-frequency solution for the total field u:

$$u = u_0 - u_n \tag{11}$$

in which $p = 0, n$,

$$u_p = \frac{2}{n}\Big\{-T_p(\eta_p)\big(2\sqrt{\pi kr}\,e^{j\pi/4}s_p(\phi)\,e^{jkrs_p^2(\phi)}h(\phi, \alpha_p) + \mathbb{F}\big[krs_p^2(\phi)\big]\big) + T_p(\eta_p) - T_p(\phi)\Big\}\frac{e^{-jkr}}{\sqrt{kr}} \tag{12}$$

where

$$T_p(\eta) = g_n(\eta)\frac{s_p(\eta)\big[s(\eta) + s_0(\phi)\big]\big[s(\eta) + s_n(\phi)\big]}{s_p(\phi)\big[s(\eta) + s_p(\phi)\big]\big[s_0(\phi) - s_n(\phi)\big]} \tag{13}$$

$$s_p(\eta) = \sqrt{2}\sin\left(\frac{\eta - \eta_p}{2}\right);\ s(\eta) = \sqrt{2}\sin\left(\frac{\eta - \phi}{2}\right) \tag{14}$$

$$h(\alpha, \alpha_p) = \begin{cases} 1, & \text{for } p = 0\ \alpha < \alpha_0,\ p = n\ \alpha > n\pi - \alpha_n \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

and $\mathbb{F}(z)$ is the UTD transition function [11] generalized to a complex argument as in [12], [13].

The compact expression (12) is applicable to any $n\pi$ wedge angle. Also it provides a clear physical description of the scattering phenomenon. It is indeed easily seen that the term multiplying $h(\alpha, \alpha_p)$ is the field of the surface wave; its discontinuity is compensated by the pertinent transition function. Thus, the field is smooth and continuous across the shadow boundaries of the surface waves.

It should be noted that when a surface wave can not be supported by the $p\pi$ face of the wedge, i.e., when $\alpha_p \le 0$, the corresponding u_p term reduces to

$$u_p = -\frac{n}{2}T_p(\phi)\frac{e^{-jkr}}{\sqrt{kr}}. \tag{16}$$

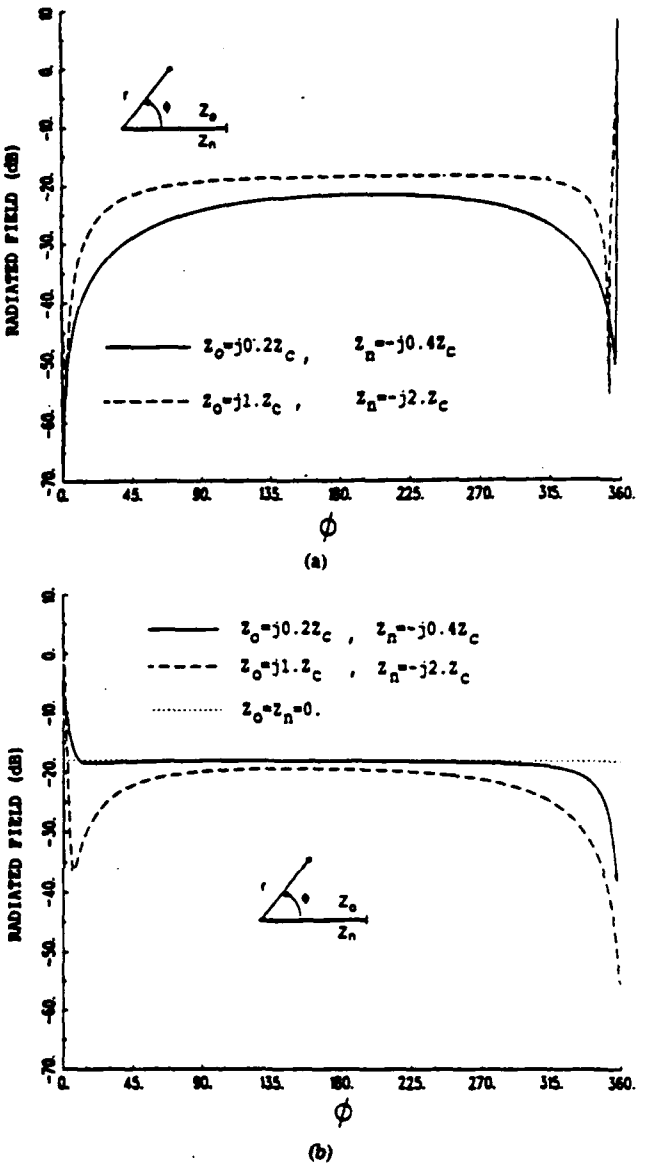


Fig. 3. Field radiated by a line source located at the edge of an impedance half-plane ($n = 2$) where $r = 10\ \lambda$. (a) Electric line source. (b) Magnetic line source.

V. Numerical Results

The solution presented in the previous section has been applied to calculate the field of both electric and magnetic line sources for several wedge angles and different impedance boundary conditions. The results obtained in some examples are presented next, to analyze the effect of the electric properties of the wedge on the excitation of the scattering mechanisms. In particular, the surface wave excitation is emphasized.

The field radiated at a distance $r = 10\lambda$ by a line source located at the edge of an impedance half-plane ($n = 2$) is plotted in Fig. 3(a) and (b).

Different purely reactive surface impedance boundary conditions (IBC's) are imposed on the faces of the half-plane. A surface wave is excited 1) on the $n\pi$ face by an electric line source (ELS) and 2) on the 0 face by a magnetic line source (MLS). Fig. 3(a) shows that the field becomes vanishingly small as the observation point approaches the faces of the wedge. However, a strong evanescent, surface wave field is excited close to the $n\pi$ face. In Fig. 3(b) the field patterns calculated for the same two different surface

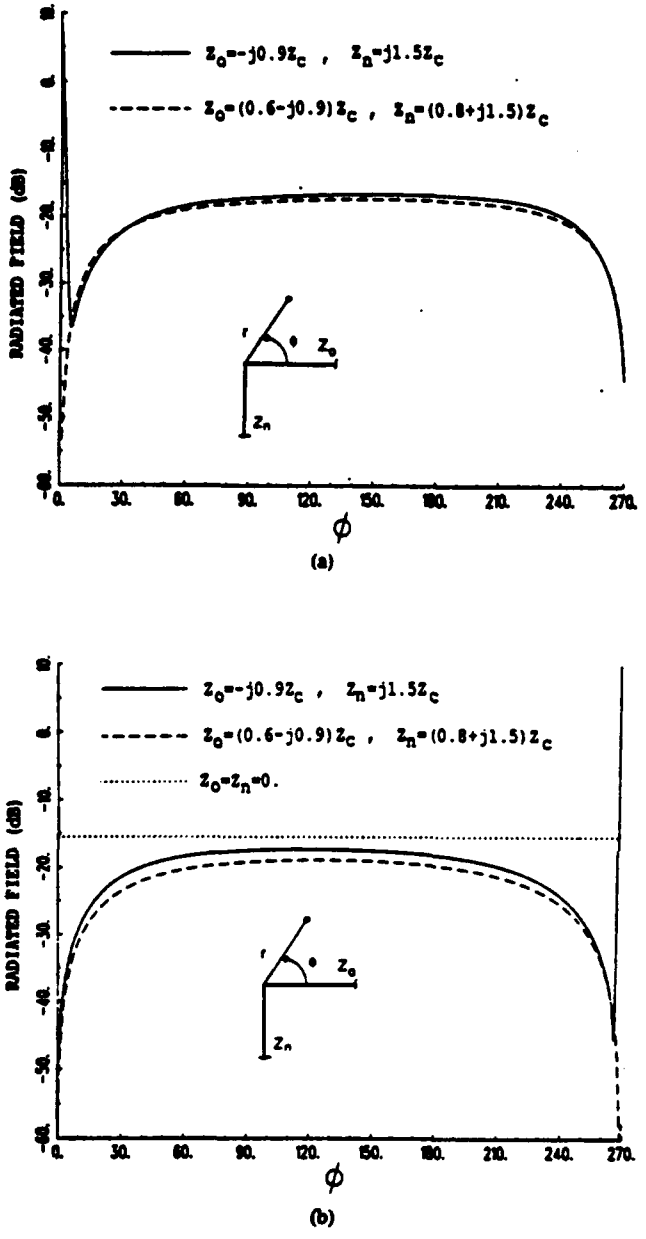


Fig. 4. Field radiated by a line source located at the edge of an impedance wedge where $n = 3/2$ and $r = 10$ λ. (a) Electric line source. (b) Magnetic line source.

impedances are compared with the reference solution for a perfectly conducting (PEC) half-plane. It is noted that at a variance of the PEC case, on one hand the field becomes vanishingly small when approaching the capacitive face, and on the other hand it is significantly enhanced by a surface wave excitation mechanism, when approaching the inductive face.

A right-angled impedance wedge ($n = 3/2$) is considered next. The fields radiated by both an ELS and an MLS are plotted in Figs. 4(a) and (b), respectively. A purely capacitive surface impedance is imposed either on the $n\pi$ (continuous line) or on the 0 face (dashed line). For an ELS (Fig. 4(a)), they cause an inductive IBC so that a strong evanescent, surface wave field contribution is observed when approaching them, as expected. On the other hand, in both cases the field decreases, more or less rapidly, when approaching the lossy inductive surface impedance faces (capacitive IBC). For an MLS Fig. 4(b) shows that introducing a loss into the inductive impedance faces chokes down evanescent field contributions. Thus, at a variance of the PEC case, the field drops down when approaching either faces of the wedge.

The above examples emphasize that in general due to the presence of the losses the evanescent field contribution is weak and observable only very close to the face. Also, it is seen that the patterns are smooth and continuous across the shadow boundaries of the surface waves.

VI. Summary

An exact integral representation for the field radiated by a line source located at the edge of an impedance wedge, has been obtained. To this end, the plane wave spectrum of the source has been used in connection with the plane wave response of the wedge. Uniform high-frequency expressions for the total field have been derived from that solution. They include the description of surface wave excitation mechanisms, and provide field patterns that are smooth and continuous across the surface wave shadow boundaries. Numerical examples have been presented, also to investigate into these evanescent wave contributions. The formulation presented here suggests a further extension to treat three-dimensional cases.

Acknowledgment

The authors wish to express their appreciation to Prof. M. Calamia for contributing useful discussions.

References

[1] R. Tiberio, G. Pelosi, G. Manara and P. H. Pathak, "High-frequency scattering from a wedge with impedance faces illuminated by a line source. Part I: Diffraction," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 212-218, Feb. 1989.

[2] W. E. Williams, "Vertex generated waves outside metallic wedges," *Proc. Cambr. Philosoph. Soc.*, vol. 57, pp. 393-400, 1961.

[3] S. N. Karp and F. C. Karal, "A new method for the determination of the far fields with applications to the problem of radiation of a line source at the tip of an absorbing wedge," *IEEE Trans. Antennas Propagat.*, vol. AP-7, pp. S91-S102, 1959.

[4] G. D. Maliuzhinets, "Excitation, reflection and emission of surface waves from a wedge with given face impedances," *Sov. Phys. Dokl.*, vol. 3, pp. 752-755, 1958.

[5] C. Gennarelli and L. Palumbo, "A uniform asymptotic expansion of a typical diffraction integral with many coalescing simple pole singularities and a first-order saddle point," *IEEE Trans. Antennas Propagat.*, vol. AP-32, pp. 1122-1124, Oct. 1984.

[6] J. L. Volakis and M. I. Herman, "A uniform asymptotic evaluation of integrals," *Proc. IEEE*, vol. 74, pp. 1043-1044, July 1986.

[7] R. G. Rojas, "Comparison between two asymptotic methods," *IEEE Trans. Antennas Propagat.*, vol. AP-35, pp. 1489-1492, Dec. 1987.

[8] G. D. Maliuzhinets, "Inversion formula for the Sommerfeld integral," *Sov. Phys. Dokl.*, vol. 3, pp. 52-56, 1958.

[9] R. G. Rojas, "Electromagnetic diffraction of an obliquely incident plane wave field by a wedge with impedance faces," *IEEE Trans. Antennas Propagat.*, vol. 36, pp. 956-970, 1988.

[10] O. M. Buyukdura and R. G. Kouyoumjian, "The radiation from scatters at the edge of a wedge," presented at the North American Radio Sci. Meeting, Vancouver, Canada, June, 17-21, 1985.

[11] R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, vol. 62, pp. 1448-1461, 1974.

[12] P. C. Clemmow, *The Plane Wave Spectrum Representation of Electromagnetic Fields*. London, U.K.: Pergamon, 1966.

[13] R. Tiberio, F. Bessi, G. Manara, and G. Pelosi, "Scattering by a strip with two face impedances at edge-on incidence," *Radio Sci.*, vol. 17, no. 5, pp. 1199-1210, Sept.-Oct. 1982.