

DTIC FILE COPY

(4)

AFGL-TR-88-0291

AD-A218 209

**A NUMERICAL AND THEORETICAL STUDY OF SEISMIC
WAVE DIFFRACTION IN COMPLEX GEOLOGIC STRUCTURE**

Gregory L. Wojcik

**WEIDLINGER ASSOCIATES
4410 El Camino Real, Suite 110
Los Altos, CA 94022**

14 April 1989

Final Report
April 1984 — July 1988

DTIC
ELECTE
FEB 13 1990
S E D
CD

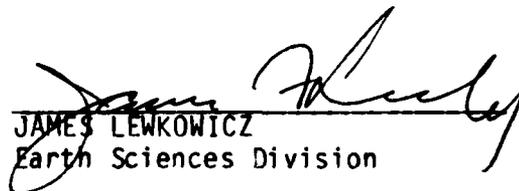
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

**AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AIR FORCE BASE, MASSACHUSETTS 01731-5000**

This technical report has been reviewed and is approved for publication.

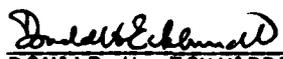


JANET C. JOHNSTON
Solid Earth Geophysics Branch



JAMES LEWKOWICZ
Earth Sciences Division

FOR THE COMMANDER



DONALD H. ECKHARDT
Division Director

This report has been reviewed by the ESD Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Center.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify GL/IMA, Hanscom AFB, MA 01731-5000. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188 Exp. Date: Jun 30, 1986	
1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS			
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited			
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) R8928		5. MONITORING ORGANIZATION REPORT NUMBER(S) GL-TR-88-0291			
6a. NAME OF PERFORMING ORGANIZATION Weidlinger Associates	6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Geophysics Laboratory			
6c. ADDRESS (City, State, and ZIP Code) 4410 El Camino Real, Suite 110 Los Altos, CA 94022		7b. ADDRESS (City, State, and ZIP Code) Hanscom Air Force Base, MA 01731-5000			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Geophysics Lab.	8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-84-C-0102			
8c. ADDRESS (City, State, and ZIP Code) Hanscom Air Force Base, MA 01731-5000		10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2309	TASK NO. G2	WORK UNIT ACCESSION NO. AR
11. TITLE (Include Security Classification) ANumerical and Theoretical Study of Seismic Wave Diffraction in Complex Geologic Structure					
12. PERSONAL AUTHOR(S) Gregory L. Wojcik					
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 4/84 TO 7/88	14. DATE OF REPORT (Year, Month, Day) 14 April 1989	15. PAGE COUNT 104		
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Seismic Modeling, Inhomogeneity, Large-Scale Simulation, Cray-2, Finite Elements, Vector Wave Diffraction Theory, Edge Diffraction, Electromagnetic Diffraction, (continued)		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>This final report presents results of a multi-year theoretical study on wave propagation and diffraction effects in irregular geologic media. The first objective of the study is to demonstrate and validate some large-scale finite element methods for analyzing linear and nonlinear seismic effects in the surficial geologies relevant to several Air Force missions. The second objective is to solve the hitherto intractable problem of vector edge diffraction, which strongly influences seismic wave interactions at the geologic discontinuities ubiquitous in near-surface structure. The presentation is in order of increasing analytical abstraction, beginning with a paper addressing various practical, numerical wave propagation issues and models, and followed by three papers on the mathematical theory of diffraction.</p> <p>The first paper describes a variety of linear and nonlinear, finite element seismic simulations at small- and large-scale, and their performance on minicomputers and super-computers, particularly the Cray-2. The verified and documented, general finite element code described has been installed on the Air Force Geophysics Laboratory's minicomputer and (cont.)</p>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Janet Johnston		22b. TELEPHONE (Include Area Code) (617) 377-3767	22c. OFFICE SYMBOL AFGL/LWH		

18. SUBJECT TERMS (continued)

Riemann-Hilbert Problems, Singular Integral Equations.

19. ABSTRACT (continued)

Air Force Weapons Laboratory's Cray-2, and is available for staff and contract researchers. Results indicate that very large 2-D models (millions of elements) are feasible, but that more speed is required for 3-D models. Nonetheless, the Cray-2 affords the opportunity to model many of the crustal experiments and propagation phenomenology of interest to the Air Force seismic community.

The second paper develops a rigorous theory of vector wave diffraction for the canonical planar wedge problem. The third paper presents a numerical evaluation and verification of the two-dimensional diffraction theory for various types of incident plane-wave polarization. The report ends with a purely mathematical paper on the theory of connected boundary value problems in multiple complex planes, which provides theoretical underpinnings for the integral representations used in the second paper, as well as more general diffraction problems. Note that the diffraction analysis is in the context of TM and TE polarized electromagnetic waves rather than seismic waves. This is relevant since the elastic SH polarization is mathematically identical to the TE (transverse electric) polarization, however, for reasons of schedule the more general P-SV-wave case is not presented. The benefit of considering electromagnetic propagation for this first published example of the solution technique and its evaluation is the broader range of application, from seismic SH-waves to light and radar waves.

In a practical vein, the diffraction solutions described here are used to verify discrete numerical solutions in the neighborhood of sharp edges. Validation is required because basis functions for the discrete solvers do not support the field singularities that actually occur at an edge. In fact, the exact solution evaluated here indicates that edge-diffracted seismic wave fields calculated by discrete numerical methods probably exhibits significant errors. (60)

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

TABLE OF CONTENTS

INTRODUCTION.....	1
MOTIVATION.....	2
DISCUSSION AND CONCLUSIONS	4

LARGE-SCALE, EXPLICIT WAVE SIMULATIONS ON THE CRAY-2

1. Introduction	7
2. Background	8
2.1 Finite Element Solvers	8
2.2 Nonlinear Constitutive Behavior.....	9
3. CRAY-2 Implementation	11
3.1 FLEX on the CRAY-2.....	11
3.2 Performance Comparisons	13
4. Large-Scale 2-D Simulations of a Maine Refraction Line	13
4.1 The Discrete Model.....	13
4.2 Finite Element Synthetic Seismograms.....	15
5. Large-Scale 3-D Simulations of a Scattering Experiment	17
5.1 The Discrete Model.....	17
5.2 Finite Element Synthetic Seismograms.....	18
6. Seismic Source Modeling	19
6.1 Elastic Source Modeling	20
6.2 Inelastic Source Modeling.....	20
7. Discussion and Conclusions	26
7.1 Large-Scale Modeling Capacity.....	27
7.2 Speed and Memory.....	27
7.3 Nonlinear Behavior, Material Attenuation, and Boundary Conditions	28
7.4 Conclusions	29

A MATHEMATICAL THEORY OF LIGHT DIFFRACTION BY DIELECTRIC WEDGES

Abstract	31
Description of the Problem	32
1. Introduction.....	32
2. Maxwell's Equations	35
3. Problem Statement.....	36
Wave Solutions and the Initial-Boundary Problem	40
4. Self-Similarity and Homogeneous Vector Solutions.....	40
5. The Vector Initial-Boundary Value Problem.....	42
6. The Characteristic Mapping.....	44
Reduction to Elliptic Boundary Conditions	46
7. Hyperbolic Solutions	46
8. Elliptic Boundary Conditions	48
Integral Representations and Singular Equations	52
9. The Scalar Riemann-Hilbert Problems	52
10. The Vector Riemann-Hilbert Problems	55
11. Analysis of Undetermined Functions.....	57
Discussion and Conclusions	60
12. Discussion.....	60
13. Conclusions.....	62
References	63

**LIGHT DIFFRACTION BY A PLANAR AIR-GLASS WEDGE:
VERIFICATION OF THE VECTOR THEORY**

Abstract.....	65
1. Introduction.....	66
2. The Singular Integral Equation.....	68
3. Solution of the Integral Equation.....	69
4. Evaluation of Diffracted Field Gradients.....	73
5. Evaluation of Diffracted Fields.....	73
6. Finite Element Model Comparisons.....	74
7. Conclusions.....	74
References.....	81

**VECTOR BOUNDARY VALUE PROBLEMS IN CONNECTED
COMPLEX HALF-PLANES**

Abstract.....	83
1. Introduction.....	84
2. Background.....	85
3. The Canonical Half-Plane Problem.....	87
4. Dual Boundary Equations and Normal Forms.....	88
5. Solutions of the Normal Forms.....	90
6. Dual Schwarz-type Integral Representations.....	92
7. Boundary Integral Equations.....	93
8. Determination of Homogeneous Solutions.....	95
9. Degenerate Solutions of the Riemann-Hilbert Form.....	96
10. Discussion.....	96
11. Conclusions.....	98
References.....	100

INTRODUCTION

This final report proffers results from a three-year theoretical study of wave propagation and diffraction effects in irregular geologic media. The results are presented in order of increasing analytical abstraction, with the first paper addressing various practical, numerical wave propagation issues and models, following by three papers on the mathematical theory of diffraction.

The ostensible purpose of the study is to demonstrate and validate some discrete numerical methods essential for analyzing linear and nonlinear seismic effects in the surficial geologies relevant to several Air Force missions. In actuality, and without apology, the study has become a pretext to solve the hitherto intractable problem of vector diffraction, which governs seismic wave interactions at the geologic edges ubiquitous in near-surface structure.

The report begins with an applications paper on the use of numerical algorithms based on explicit finite element methods. It describes a variety of linear and nonlinear seismic simulations at small- and large-scale, and their performance on minicomputers and supercomputers. The verified and documented, general finite element code described in the paper has been installed on the Air Force Geophysics Laboratory's minicomputer and is available for staff and contract researchers. Various aspects of finite element seismic modeling incorporated in the code have been supported by AFGL over the years and the benefits have now come full circle.

The second paper develops a rigorous theory of vector wave diffraction for the canonical planar wedge problem. The third paper presents a numerical evaluation and verification of the two-dimensional diffraction theory for various types of incident plane-wave polarization. The report ends with a purely mathematical paper on the theory of connected boundary value problems in multiple complex planes, which provides theoretical underpinnings for the integral representations used in the second paper, as well as more general diffraction problems.

Note that the diffraction analysis is in the context of TM and TE polarized electromagnetic waves rather than seismic waves. This context is relevant since the elastic SH polarization is mathematically identical to the TE (transverse electric) polarization. For reasons of schedule, the more general P-SV-wave case is not presented, although it involves relatively minor modification of the electromagnetic theory. The benefit of considering electromagnetic propagation for this first published example of the solution technique and its evaluation is the broader range of application, from seismic SH-waves to light and radar waves.

In a practical vein, the diffraction solutions described here are used to verify discrete numerical solutions in the neighborhood of sharp edges. Validation is required because basis functions for the discrete solvers do not support the field singularities that actually occur at an edge. In fact, the exact solution evaluated here indicates that edge-diffracted seismic wave fields calculated by discrete numerical methods probably exhibits significant errors.

MOTIVATION

Stimulus for this research is the patent observation that near-surface geology can deviate markedly from the seismologist's common idealization of a uniformly layered half-space. This idealization has been assumed almost universally in quantitative seismic analysis and interpretation until the last fifteen years or so, as more attention has been focused on near-field propagation of high frequency waves. These deviations from uniform layering are caused by tectonics, sedimentation, and weathering—exhibited as folds, faults, escarpments, basins, mountains, etc., often two-dimensional and over a range of length scales. Consider the illustrations in Fig. 1 for some relevant examples. Clearly, this complexity must be incorporated in any mathematical model used to understand near-surface seismic wave interaction. Such interaction is important in many practical applications, for example, ground motion estimates near a strategic site, the modification of source radiation pattern, vehicle or event detectability and location uncertainty, the possibility of seismic shielding, general scattering and attenuation, and so on. Such models are routinely used for critical or strategic site analysis as well as nuclear explosion detection and discrimination.

Often, by virtue of these complicated surficial geologies, the only recourse for modeling seismic effects is numerical simulation—typically using geometric or discrete methods. Geometric methods constitute ray tracing and its various extensions. Although they generally involve high frequency asymptotics and yield incomplete solutions, they are certainly useful in their place. Discrete methods like time-domain finite differences (FDTD) and finite elements (FETD) provide complete, broadband solutions but require substantial computer resources. They also limit the analyst's ability to generalize results without an expensive suite of calculations. Despite these limitations, discrete methods are becoming more and more prevalent, particularly with new, more robust algorithms and the availability of large-memory supercomputers. Although much more complete than geometric methods, discrete methods have their own shortcomings, e.g., low and high frequency cutoffs, dispersion, model truncation, low order elemental approximations, etc.

Clearly, regardless of the analytical method used, the model's numerical limitations must be well-established before making predictions or basing critical decisions on its predictions. One very prominent limitation of both geometrical and discrete methods that has never been adequately addressed is the fidelity of diffractions from model edges. The edge is a linear feature (lineament) of high curvature on an interface separating different geologic media, and introduces elastic field singularities and diffracted radiation that are not included completely in any ray tracing or discrete solver formulation existing today. The reason is that the canonical mathematical problem of elastic edge diffraction has never been solved, hence, cannot be incorporated. Sharp edges, i.e., sharp with respect to the wave length, are ubiquitous in near-surface geology, e.g., faults, basin edges, and escarpments. Therefore, the seismic community needs some quantification of the phenomenon. This issue is the principal motivator for the research reported herein.

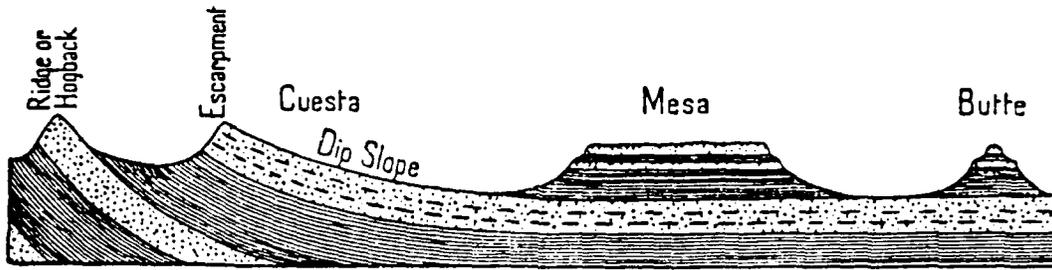


FIG. 413 Diagram to illustrate the relation of various erosional landforms to the structure and dip of the strata from which they have been carved

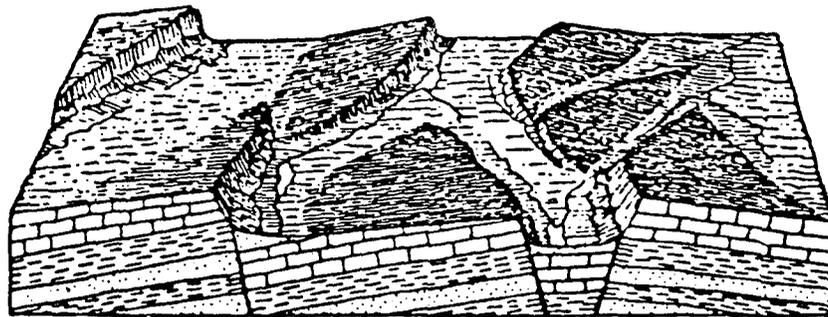


FIG. 796 Diagram to illustrate the fault-block structure of the ranges of the Great Basin, Utah (After W. M. Davis)

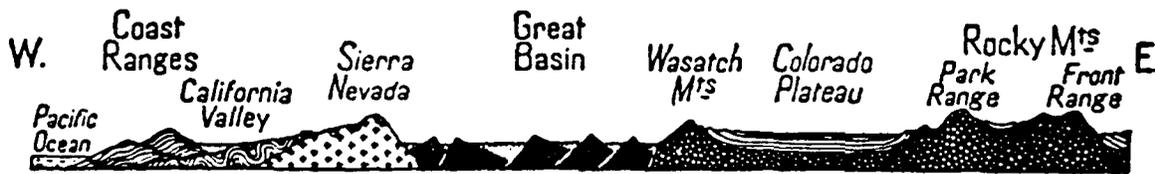


FIG. 791 Schematic section across the Cordillera of the western United States. Length of section about 1,900 miles (3,000 km.)

Figure 1. Examples of erosional and tectonic landforms with edged features promoting diffraction and scattering, reproduced from Holmes (1965).

DISCUSSION AND CONCLUSIONS

The first paper on applications demonstrates our present ability to simulate large-scale, time-domain wave propagation in inhomogeneous or nonlinear models using explicit, finite element techniques. The examples show that one of the most powerful supercomputers available today, the Cray-2, can indeed perform very large, two-dimensional, elastic simulations in complex models very effectively. This computational resource affords an opportunity for analysts to model, numerically, many of the crustal experiments and propagation phenomenology issues of interest to the Air Force seismic community. Although these explicit, linear wave problems are well-solved by virtue of the high degree of vectorization they admit, the addition of nonlinear constitutive algorithms markedly reduces performance since they do not vectorize in general.

The calculations also indicate that, for three-dimensional problems commonly encountered in applied seismology, still more speed is required. This is not to say that useful simulations cannot be done, e.g., the scattering analysis discussed in the paper shows that some limited but interesting three-dimensional simulation can finally be addressed at reasonable cost. What must be emphasized is that cost grows so quickly in the third dimension that only processor speed increases by one, or better yet, two orders of magnitude can push affordable model size beyond the minimum required for adequate temporal and spatial resolution. Multiprocessor hardware with distributed algorithms, e.g., Connection type machines, or vector-parallel machines with suitable compilers, e.g., Alliant or modern hypercube type, provide another avenue to achieving the necessary speed increase.

Simulations of the type described in the first paper are the only means of investigating propagation phenomena involving range, azimuth, and depth dependent, i.e., nonseparable, or nonlinear model properties. Other options like ray tracing or boundary integrals involve idealizations that are too often unacceptable for this broader class of earth models. Note that discrete simulation has, here and elsewhere, been used principally for the forward problem. As the above performance issues are addressed more fully by software and hardware, it will eventually become possible to implement discrete models for the inverse problem, i.e., in an optimization loop with the synthetic data generated from discrete trial models. This possibility is perhaps the ultimate goal of seismic simulation research.

One obvious limitation of the models and simulation described above, is our inability to verify certain aspects of the scattered wave field. This issue is suggested by the two-dimensional refraction model in the first paper, where strong conversions of seismic radiation to Rayleigh waves is observed at points of model surface roughness, e.g., at a low mountain, and the discontinuities at basin edges. The question is, to what extent is this a numerical artifact of the finite element discretization? The only way to answer this question is to compare against canonical solutions around edges in so-called benchmark problems. The diffraction theory developed and

verified in the second, third, and fourth papers provides such a benchmark. In fact, the verification comparisons reported in the third paper indicate that a careful examination and reappraisal of any discrete numerical solution near sharp edges or rough structure is warranted.

To this end, the full P-SV elastic diffraction theory needs to be derived and an evaluation performed, comparable to the work reported here in the second and third papers on electromagnetic diffraction. Only from such a solution can the various algorithms currently used today be verified for application to problems involving rough topography or other strongly diffracting structure. Furthermore, these analytical solutions can provide a mathematical basis for incorporating enhanced diffraction capabilities in both discrete and geometric modeling algorithms.

In terms of further extension of the mathematical theory presented here, the vector wave diffraction solution was derived by intuition, generalization of known mathematical techniques, and persistence—not by conceptual leaps or new mathematics. The same intuition and mathematical generalization provide a basis for extending the two-dimensional approach to three-dimensions, applicable to a variety of diffraction problems found in radar and optical imaging applications as well as stealth technology. Clearly, for these reasons as well as seismic benchmarking and algorithm enhancement, this type of theoretical research should be continued.

In conclusion, the research contract under which this work was performed has produced a number of substantial benefits and returns, specifically for the Air Force Geophysics Laboratory, but also for other military and industry groups. These are both practical and intellectual.

The practical benefit is availability of a general purpose seismic wave modeling capability—in the guise of a fully documented, verified, and demonstrated finite element code for one-, two-, and three-dimensional models—to AFGL researchers and contractors. It runs effectively on machines ranging from PCs to the most powerful Cray supercomputers. However, it is *one thing* to develop a comprehensive, rigorous seismic modeling code, and quite another to apply it constructively to many of the seismological problems requiring such capabilities. Making it available to this community should substantially increase the knowledge return to the Air Force, particularly in view of the supercomputer resources now available on Air Force systems.

The intellectual benefit is that, as a result of this research, the most intractable diffraction problem in the classical theory of wave propagation has now been solved. This is significant because of its intrinsic mathematical importance and the fact that it has confounded analysts for the last century. The solution governs electromagnetic, elastic, and acoustic wave propagation in the neighborhood of a planar edge and consequently has a wide variety of non-seismic applications. In the seismic context, its practical implications are in terms of augmenting discrete modeling algorithms to better represent diffraction phenomena.

LARGE-SCALE, EXPLICIT WAVE SIMULATIONS ON THE CRAY-2

G.L. WOJCIK, D.K. VAUGHAN, M. BARENBERG and J. MOULD

Weidlinger Associates, 620 Hansen Way, Palo Alto, CA 94304, U.S.A.

M.B. HULIT

Air Force Geophysics Laboratory, Hanscom Air Force Base, MA 01731, U.S.A.

Most time-domain, wave modeling problems in geophysics are intractable by classical analysis methods, due principally to nonseparability and to a lesser extent material nonlinearity. Therefore discrete numerical solutions are often necessary for the simulation of realistic models. Applications in 2-D and 3-D geophysical modeling are the subject of this paper, particularly as solved on a CRAY-2 supercomputer. Implementation and performance differences between earlier CRAYs and the CRAY-2 are described, including the discrepancy between scalar fetch and vector processing speeds. Explicit finite element solvers are applied to applications involving 2-D simulation of a seismic refraction experiment across the state of Maine, 3-D simulation of near-source scattering experiments, and both linear and nonlinear axisymmetric source simulation. Results show that the CRAY-2 allows cost-effective 2-D simulations of truly large-scale models, but only begins to be effective in 3-D for models of interest in geophysics. The large memory (256 megawords) is adequate but a speed increase of at least an order of magnitude is necessary for cost-effective 3-D. True multiprocessor capability (rather than 'multi-computer') provides an alternative to raw speed but involves another set of difficulties.

1. Introduction

A large number of the time-domain, wave modeling problems in geophysics are intractable by classical analysis methods—by virtue of either nonseparability or nonlinearity. In fact, only a few practical problems are addressed by classical analysis, i.e., separable and linear, although this restricted class has received most of the theoretical attention. The broader class of nonseparable, nonlinear problems requires discrete numerical solutions. Some of these discrete time-domain wave propagation problems in geophysics are the subject of this paper, particularly as they are implemented and solved on the CRAY-2 supercomputer. The paper emphasizes various performance and modeling issues, with little attention given to analytical development.

By way of background recall that separability depends on the medium's degree of inhomogeneity, namely, whether it conforms to a separable coordinate system for the governing partial differential equations. Separability is not a good global assumption in general geophysical applications. In contrast, the typical assumption of linearity depends on the medium's constitutive model and, excluding the immediate source region, is often a good global assumption.

Separable problems include the common homogeneous or vertically stratified half-space and are well solved by classical methods, with weak inhomogeneities sometimes included via perturbation methods. For truly nonseparable problems, characterized by significant deviation from the layered half-space, no corresponding analytical methods are available. Either local

geometrical solutions or global numerical solutions are necessary, provided in practice by geometrical diffraction and ray theory or discrete numerical methods like finite difference and finite element wave solvers, and to a lesser extent by boundary integral methods.

Nonlinear continuum mechanics problems typically involve irreversible material behavior due to strains beyond the elastic limit (constitutive). Nonlinearities may also be due to large displacements and/or rotations (geometric). However, the contribution of geometric nonlinearities is generally secondary to constitutive effects in geophysical modeling. An effective means of including these constitutive nonlinearities is through a rigorous plasticity formulation.

The 'best' approach for practical global solutions of nonlinear, nonseparable wave propagation problems in geophysics is a discrete method. Reasons for this choice include ease of modeling, minimal need for geometric or material idealization, full wave representation, and the availability of efficient algorithms in conjunction with supercomputers. Certainly, discrete methods have their share of drawbacks as well, including the inability to generalize results of one calculation beyond its basic parameters, the difficulty of separating wave phenomena, e.g., body waves and surface waves, and errors associated with a finite model boundaries.

To address both the pros and cons of discrete numerical methods in geophysics, this paper describes some applications of explicit finite element solvers to large-scale wave modeling problems, principally on the CRAY-2 supercomputer. These include 2-D modeling of a refraction experiment across the state of Maine, 3-D modeling of some near-source scattering experiments, and both linear and nonlinear source modeling simulations.

2. Background

The discrete wave solvers applied here incorporate finite element reductions of the governing partial differential equations to ordinary differential equations in time. These ODEs are integrated forward in time using a modified leapfrog scheme (centered differences). The routines are included in a pre- and post-processing shell called FLEX, designed for efficient modeling, solution, and interpretation of large-scale propagation problems (mechanical or electromagnetic) [1]. The code was originally written to take advantage of the architecture and power of CRAY computers by developing fully vectorized kernel processing loops first, followed by an efficient code architecture to support them. Rather than use one general purpose processing routine, a group of specialized finite element and finite difference routines was developed for each class of problems, including 1-D, plane and axisymmetric 2-D, and full 3-D. FLEX was designed to minimize storage requirements for each problem by using an internal data management system that avoids the use of dimensioned arrays within the code. The system automatically and adaptively sets up internal data arrays as required by the problem at hand. Very large problems are efficiently stored in memory, thus avoiding I/O limitations of data transfer to disk.

2.1. Finite element solvers

The explicit finite element routines applied in this paper to linear and nonlinear continuum mechanics problems employ bilinear (2-D) or trilinear (3-D) shape or interpolation functions over each element (e.g. see [2]). Element geometry is either Cartesian (rectangles and bricks in 2-D and 3-D, respectively) or skewed. The term, explicit (in contrast to implicit), effectively means

that lumped masses are employed. This eliminates the need for assembly of a global stiffness matrix and a mass matrix inversion. Instead, nodal forces are accumulated element-by-element and the nodal (lumped mass) velocity increments are calculated from Newton's law in incremental form. The algorithm is stable provided the timestep is less than the fastest wave's transit time across the smallest element. This follows since all nodes are effectively decoupled by the fundamental hyperbolicity of the governing equations; in other words, the influence of a node during one timestep cannot affect nodes more than one element away. Numerical noise (roundoff) propagates one element per timestep.

The algorithm stores three types of quantities for each node—lumped mass $M(n)$ velocity vector $V(n)$, and force vector $F(n)$, where $n = 1, \dots, N$, with N the total number of nodes. The nodal velocity and force vectors' dimension, d , is equal to the problem dimension, so there are seven quantities per node in 3-D and five in 2-D (plane or axisymmetric). Velocities are defined at full timesteps while forces are defined at timestep midpoints. Nodal displacements ($U(n)$) defined at timestep midpoints are not calculated directly, but may be found by integrating velocities. Similarly, stresses are not explicitly calculated, but if required, e.g., in nonlinear calculations, they are defined at the element centroid at timestep midpoints and updated using incremental displacements.

The basic algorithm includes a vectorized force loop over rows of physically contiguous elements (row by row) and a velocity loop over all of the nodes. In the force loop, increments ($\Delta F = K\Delta U = KV\Delta t$) are calculated at the nodes of each element and added to the nodal force vector to obtain forces at the next half timestep ($F = F + \Delta F$). After all of the elements are processed a single nodal loop updates velocities. In this loop, velocity increments ($\Delta V = M^{-1}F\Delta t$, where M is diagonal) are calculated using the nodal forces at the previous half timestep, and added to the nodal velocity vector to obtain values at the next full timestep ($V = V + \Delta V$). These loops are repeated for the required number of timesteps.

Ninety percent of the floating-point operations in the algorithm occur in the element force loop. For example, in a 2-D plane, elastic model, this loop fetches nodal velocities and forces as well as material and shape information, performs 73 floating-point operations, and stores updated nodal forces for each element in the row. In general, data arrays for the nodal and elemental quantities are mapped into memory so that contiguous elements in a row have contiguous storage locations. The resulting data structures are such that a row of elements of any length can be integrated one timestep in a vectorized computation without the need of a gather operation to fetch data into a contiguous local array, and a scatter operation on the results.

The code lends itself to efficient vectorization in 'vanilla' FORTRAN (i.e., no assembly coding necessary), and approaches the peak performance levels expected for 'nonideal' problems, i.e., inhomogeneous, on pipelined supercomputers like the CRAY machines. This is not to say that a supercomputer is required for practical calculations: tens of thousands of elements are routinely executed on minicomputers, up to perhaps 100,000 elements, with reasonable execution times.

2.2. Nonlinear constitutive behavior

Nonlinear material behavior can be modeled by a variety of constitutive relations and numerical implementations. For soil-type and rock-type continua the so-called cap model has proven effective, particularly in the context of explicit finite element or finite difference wave

solvers. The cap model is essentially an algorithm implementing a rate-independent plasticity theory with an associated flow rule. Since the formulation is not as well known as the discrete wave propagation algorithms utilized in this paper, it is reviewed below.

The cap model, e.g., [3], is based on classical plasticity theory with incremental stress-strain relations in the form

$$\dot{\sigma} = C\dot{\epsilon}, \quad (2.1)$$

where σ is the stress tensor (tension positive), ϵ the strain tensor, and C the (tangent) modulus matrix, assumed positive semi-definite to insure uniqueness, stability, and continuity. The cap model is defined by a convex yield surface, $Y(\sigma)$, and a plastic strain rate vector, $\dot{\epsilon}^P$, which is normal to the yield surface in stress space so that

$$\dot{\epsilon}_{ij}^P = \lambda \frac{\partial Y}{\partial \sigma_{ij}}, \quad (2.2)$$

where λ is a non-negative scalar function.

Bulk modulus K and shear modulus G are used to represent simple linear elastic behavior inside the yield surface. The surface is defined by a fixed failure envelope and a hardening cap. The failure envelope is defined by

$$Y(\sigma) = \sqrt{J_2'} - F_F(J_1), \quad (2.3)$$

$$F_F(J_1) = A - C e^{BJ_1}, \quad (2.4)$$

where A , B , and C are material parameters, J_1 is the first invariant of the stress tensor, and J_2' the second invariant of the deviatoric stress tensor. The cap is defined by

$$Y(\sigma) = \sqrt{J_2'} - F_c(J_1, \kappa) \quad \text{for } L(\kappa) \geq J_1 \geq X(\kappa), \quad (2.5)$$

where κ is an internal state variable that measures hardening as a functional of plastic volumetric strain history, and $L(\kappa)$, $X(\kappa)$ define the J_1 -range of the cap. Function $F_c(J_1, \kappa)$ is defined by

$$F_c(J_1, \kappa) = \frac{1}{R} \sqrt{[X(\kappa) - L(\kappa)]^2 - [J_1 - L(\kappa)]^2}, \quad (2.6)$$

in which

$$L(\kappa) = \begin{cases} \kappa & \text{if } \kappa < 0, \\ 0 & \text{if } \kappa > 0, \end{cases} \quad (2.7)$$

$$X(\kappa) = \kappa - RF_F(\kappa), \quad (2.8)$$

where R is a material parameter.

Hardening parameter κ is implicitly defined as a functional of the plastic volumetric strain, ϵ_v^P , by means of a relation between $X(k)$ and ϵ_v^P , which is then coupled with (2.8) to define k in terms of ϵ_v^P . For soils, the relationship is

$$\tilde{\epsilon}_v^P = W[e^{D(X(\kappa) - X_0)} - 1], \quad (2.9)$$

in which W , D , and X_0 are material parameters and ϵ_v^p is a history-dependent functional of the plastic volumetric strain and is given by the differential functional relation

$$\dot{\epsilon}_v^p = \begin{cases} \dot{\epsilon}_v^p & \text{if } \dot{\epsilon}_v^p \leq 0 \text{ or } \kappa < 0, \\ 0 & \text{if } \dot{\epsilon}_v^p > 0 \text{ and } \kappa \geq 0. \end{cases} \quad (2.10)$$

If dilatancy occurs when $J_1 > 0$, (2.10) limits the shrinking of the cap to $\kappa = L(\kappa) = 0$. This insures that the cap remains finite and is assumed to apply in general for soils. Such a limitation is somewhat arbitrary in view of the lack of data with respect to material behavior after the occurrence of tension failure.

Tensile stresses are limited by the condition that the first invariant of stress

$$J_1 \leq T, \quad (2.11)$$

so that T (a material parameter) represents a tension cutoff parameter. In order to insure that no principal stress exceeds parameter T , the slope of the failure envelope (a measure of the friction angle) has been appropriately limited.

The cap model provides a theoretically sound idealization for the salient features of nonlinear soil behavior. The solution algorithm [3] is robust and quite efficient relative to many other nonlinear material models and has proven effective for a wide variety of applications.

3. CRAY-2 implementation

The CRAY-2 was the largest and fastest 'conventional' supercomputer available at the time the calculations described in this paper were performed (mid 1987). It uses a UNIX-based operating system, has four available central processing units, and addresses 256 million words of CMOS memory (typically). Other machines are available with theoretically faster architectures, e.g., the Connection Machine, but they also require significant modification of the discrete algorithms and are presently unproven for production problems.

The CRAY-2 is therefore the machine of choice for large-scale simulations, particularly since similar CRAY-1 type machines have been available for years and optimal programming techniques are fairly well understood. Note that the principal advantage of the CRAY-2 over earlier CRAYs (the X-MP for example) is the large memory available. The factor of two to three in speed is also significant but not the principal reason for its choice. The availability of four processors is not very useful in practice since chargeable time is accumulated on each one, with nontrivial overhead for multiprocessor usage. The only advantage is a factor of three or so decrease in wall clock time. This is useful for reducing run time, given the possibility of a system crash during extended runs. A better solution is to store memory images periodically for restarts during long runs (> five hours or so).

3.1. FLEX on the CRAY-2

Several differences were found between the performance of FLEX on the CRAY-2 and its performance on earlier CRAY machines. First, the original CRAY-1 FORTRAN coding produced erroneous results on the CRAY-2 caused by the use of CDIR\$ IVDEP system

commands that force vectorization of certain computational loops containing vector dependencies. Specifically, this occurs for the elemental loops computing internal force contributions to each node of an element. Since a node receives force contributions from two adjacent elements, it is clear that if the internal force contributions of both elements are being computed during the same pass through the loop, there is a potential dependency problem. This issue was considered during the initial development of FLEX and it was determined that forced vectorization gave correct results. However, this did not carry over to the CRAY-2. The problem was easily corrected by processing each row of elements in two passes, using a do loop increment of 2 to ensure that two adjacent elements having common nodes would not be processed during the same pass through the computational loop. This guaranteed that no vector dependencies would occur if vectorization was forced.

A second difference worth noting between the CRAY-2 and earlier CRAYs is the speed of scalar fetches from scattered memory locations. This issue involves the tradeoff between increased storage versus CPU time in the design of efficient processing strategies, in particular, to optimize storage for very large problems. For machines with 2-4 million words of storage, it was considered wasteful to store often duplicated material parameters for each element in the model, i.e., bulk and shear moduli. It even seemed extravagant to store a single material number of each element, since for a model containing one million elements, there are often less than 20 unique sets of material properties. The approach adopted on earlier CRAYs was to develop a group of unique site profiles, condensed to minimum storage, and a pointer table defining the site profile for each row of elements. In this way material properties of a homogeneous 1000×1000 element model can be completely defined by 1005 words of storage compared with 1,000,002 words if storing the material model number for each element and 2,000,000 words if storing the bulk and shear moduli for each element.

When computing the internal resisting forces for a row of elements, the site profile for the row is checked against the previous row's profile, and if they differ, the condensed site profile for the new row is expanded and loaded, with one word for each element. The element material properties are loaded into the bulk and shear moduli arrays in a scalar loop since this process does not vectorize on the CRAY-2. On earlier CRAYs, the overhead associated with retrieving the two material parameters at arbitrary locations in memory and placing them in consecutive locations in a local array (to facilitate vectorization) was a factor of two in the worst case, i.e., 2 times the homogeneous site performance. Although expensive, this approach significantly increased the size of models resident in memory, thus eliminating expensive disk I/O. In contrast, on the CRAY-2 it was found to increase the overall processing cost by at least a factor of 3.3.

The increased overhead follows because scalar fetches on the CRAY-2 are only slightly faster than on the CRAY-1, but since vector processing is about 2.6 times faster the scalar overhead is more significant. This can decrease the overall performance of the CRAY-2 to below CRAY-1 levels. After review, system analysts considered this scalar performance as normal for the CRAY-2's memory architecture and recommended replacing the single scalar loop fetching material properties with a series of gather calls. The result of this effort was to slow processing another 10 percent.

Based on these findings, it is clear that savings in storage afforded by using the site profile approach are irrelevant on the CRAY-2 with its large memory, in view of the significant CPU overhead incurred by scalar fetches of material properties. Consequently, the CRAY-2 version of FLEX has been modified to store both the bulk and shear moduli for each element in the model.

3.2. Performance comparisons

With the changes described above, FLEX achieves performance levels on the CRAY-2 that are about 2.6 times those on the CRAY-1 and 1.8 times the CRAY X-MP. These comparisons are for the same model and site profile description. For 2-D elastic axisymmetric problems, FLEX can compute 1.4 million element-timesteps per second of CRAY-2 CPU time, and for 3-D problems, it computes 0.28 million element-timesteps per second. For example, integrating a 1000×1000 element 2-D problem 2000 timesteps (2 billion element-timesteps) requires $2 \times 10^9 / 1.4 \times 10^6 \approx 1429$ seconds ≈ 24 minutes, and integrating a $100 \times 100 \times 100$ element 3-D problem 200 timesteps (200 million element-timesteps) requires $2 \times 10^8 / 0.28 \times 10^6 \approx 714$ seconds ≈ 12 minutes.

The major advantage of the CRAY-2 is of course its large memory. For the explicit finite element algorithm used in FLEX, memory requirements are 5 and 7 numbers per node in 2-D and 3-D respectively. Storage of synthetic seismograms typically adds another 10 to 30 percent to this, depending on the number of outputs points and components. Thus, the million element 2-D and 3-D examples cited above would require on the order of 6 and 9 million words respectively for the model and output. The calculations described below routinely use 10–30 million words, with no disk access. No degradation in performance is observed as memory usage increases, which has been verified for problems accessing up to 200 million words.

4. Large-scale 2-D simulations of a Maine refraction line

The first example calculation models a 200 km refraction line shot along an approximate southeasterly track across the state of Maine. This was the longest line instrumented by the United States Geological Survey during a 1984 cooperative refraction experiment in the north-eastern United States and southeastern Canada, e.g., see [4]. The line is perpendicular to the dominant structural features in the area and is reasonably approximated by a 2-D model. There were three large, high explosive shots recorded on the line—at each end and in the middle—but only the northwestern shot is modeled here.

4.1. The discrete model

The geologic model is illustrated in Fig. 1, showing a 50×183 km section of the Earth's crust and upper mantle with piecewise homogeneous structure—based on USGS data and interpretations [5]. Near-surface seismic velocity features are represented in some detail in the section, as well as the dominant topographic features (although not apparent in the figure). The finite element model was composed of a uniform 1000×3660 Cartesian mesh (i.e., non-skewed) of 50×50 meter elements, for a total of 3.66 million elements (7.33 million nodal equations). Interfaces and free surface topography were resolved step-wise by the mesh as in conventional finite difference discretizations. Clearly, the CRAY-2 is essential for a model of this size since the entire calculation (model and data) resides in-core to eliminate costly memory swaps to disk.

Although the geology is primarily 2-D on the Maine refraction line, the use of an effective point source clearly makes the problem 3-D. Since 3-D modeling on the scale of this experiment is impractical on any modern supercomputer (requiring on the order of 1000 times more power

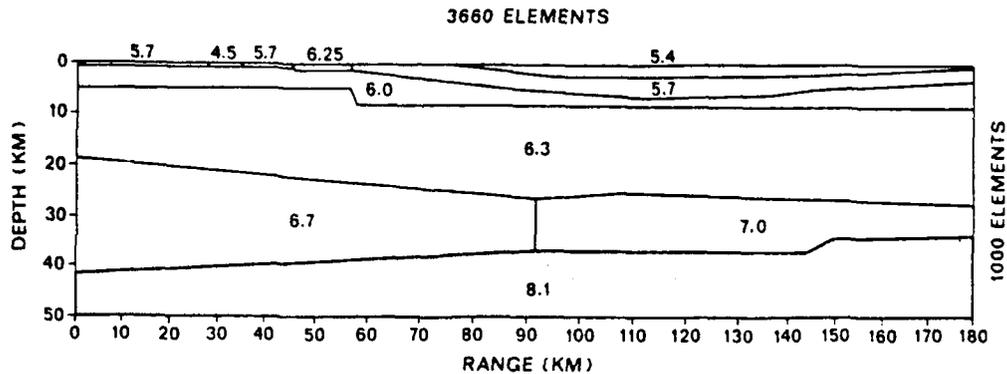


Fig. 1. A geologic model of the Earth's crust and upper mantle on a southwest section across the state of Maine, from [5]. P-wave speeds are indicated in each homogeneous 'layer'. Finer near-surface velocity structure and topography are included but not apparent in the drawing.

and memory than presently available), a 2-D approximation is necessary. This was accomplished by posing the problem as axisymmetric, with the source on the axis of symmetry, i.e., the left (northwest) boundary of the model. Axial symmetry is preferable to a plane strain approximation because it exhibits the proper radial divergence, however, the resulting global model in 3-D, obtained by sweeping the geologic section about the axis, certainly appears unphysical.

Boundary conditions applied to the finite element model include symmetry on the left side as described above, a so-called absorbing or radiation condition on the bottom and right, and a free-surface condition on the top. The absorbing boundary condition used here and in subsequent finite element calculations described below is the simple normal impedance condition based on the unidirectional wave solution, $[\sigma] = -\rho c[v]$, relating jumps in stress and particle velocity. This is implemented for both P- and S-wave incidence and does a surprisingly good job for non-normally incident waves, particularly the P-wave. It is also readily vectorizable.

The seismic source used in the USGS refraction experiments was a 2000 pound cylindrical charge of chemical high explosive. The six inch borehole was 150 feet deep, filled to a depth of 75 feet with explosive, and topped off with rained sand. Considering that the entire borehole is contained in a single element of the large-scale finite element model, it was not practical to include details of the source directly. Instead a simple surface pressure, actually lumped vertical forces, with Gaussian distribution over a few nodes was applied to excite seismic waves. No attempt was made to reproduce the actual source's radiation pattern over takeoff angle. Linear and nonlinear source simulations will be considered further in a subsequent section.

The source time function is chosen to match the observed or expected frequency content at some range—in order to better compare synthetics to actual seismograms. However, for linear models it is often more useful to calculate the impulse or step response, from which synthetic seismograms due to an arbitrary source time function are obtained by convolution. The frequency content of the derived response is of course limited by discretization, which, for the low-order elements (linear interpolation) used here, typically means significant dispersion when a wavelength is supported by eight elements or less. This behavior is implicit in the step or impulse response.

4.2. Finite element synthetic seismograms

Calculated response of the finite element model is illustrated in Figs. 2 and 3, showing synthetic vertical velocity seismograms at two kilometer station increments between 0 and 180 km from the northwestern source. They show true ground motion, i.e., no instruments response is included. Step response is plotted in the bottom suite of seismograms, and after convolution with 2 Hz and 0.5 Hz wavelets in the middle and top suites respectively. The synthetics in Fig. 2 are scaled by range (multiplied by radius) to remove spherical spreading attenuation, and further multiplied by ten to accentuate early arriving body waves. The synthetics in Fig. 3 are scaled by the square root of range to remove the cylindrical spreading attenuation of surface waves and clearly show the Rayleigh wave.

The seismograms in Fig. 2 show a variety of body wave behavior, including reflection and refraction arrivals from the intermediate layers (P_g), reflections from the deeper interfaces (P^* and P_mP), mode conversion of Moho reflected body waves (P_mP) to Rayleigh waves by

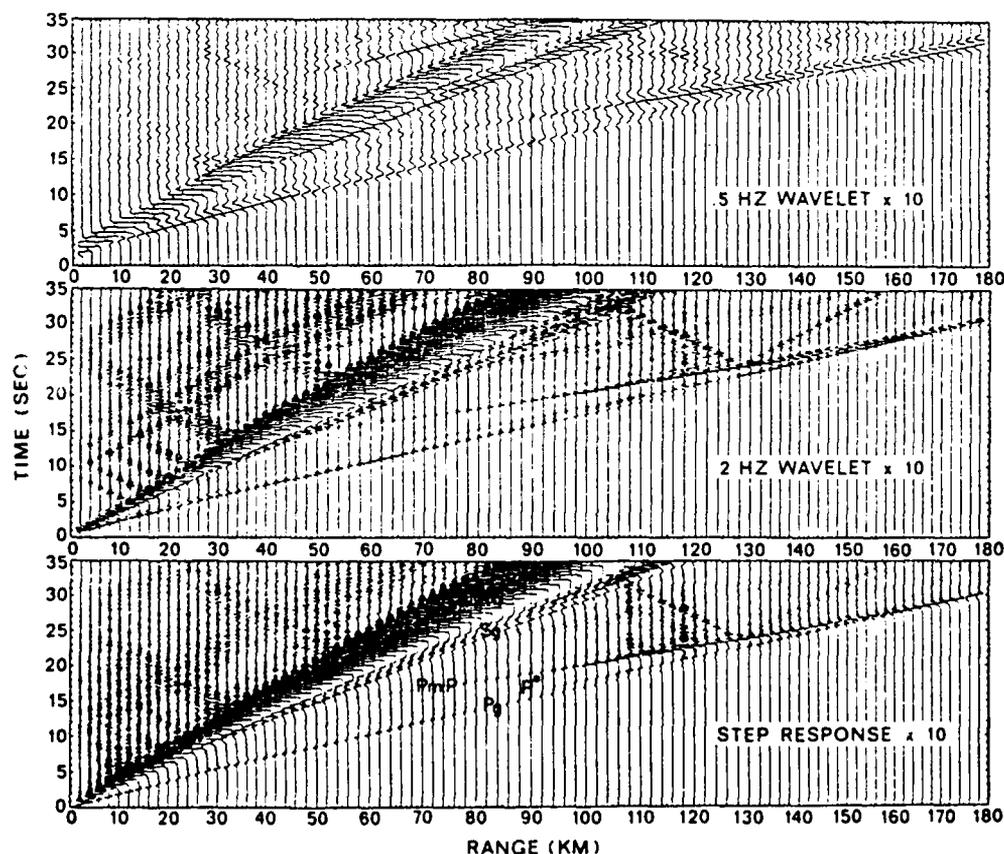


Fig. 2. Suites of magnified vertical velocity seismograms at 2 km station spacing across the Maine model. The records clearly separate body and surface wave phases, which are labeled in the lower suite. The wavelet responses are obtained from the step response by convolution.

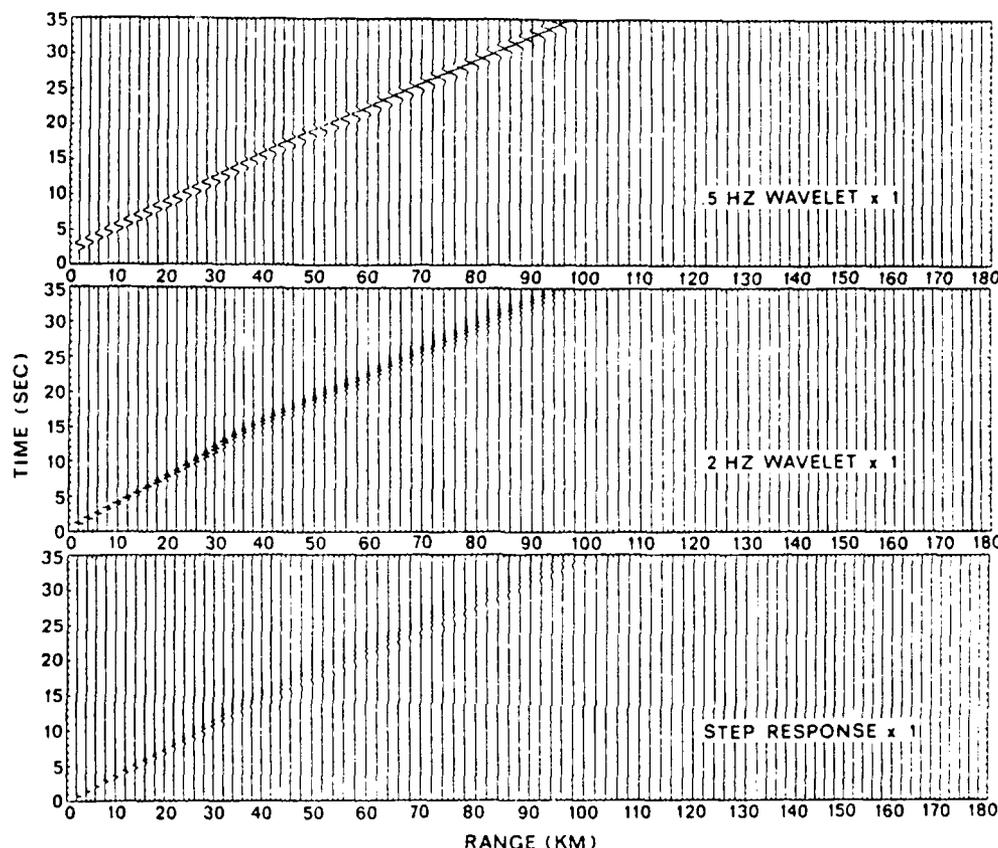


Fig. 3. Suites of unmagnified vertical velocity seismograms across the Maine model, showing predominance of the Rayleigh surface wave (R_g).

topography, and layer resonance of Moho reflections at distinct stations. They also show clear Rayleigh to Rayleigh reflections from lateral velocity variations in the top layer, as well as from topography. Observe that the various coherent phases are readily identified from the suites of seismograms by their slope, i.e., inverse phase velocity. These have been measured from the 2 Hz suite in Fig. 2 for the following interpretations.

The first arrivals out to about 110 km are P-wave reflections and refractions from the first interface, while from 70 km and beyond the arrivals transition into deeper reflections and refractions. Corresponding S-wave arrivals can also be seen preceding the Rayleigh wave, which is clipped in these synthetics. Discernible from 50 km (subcritical) and stronger after 100 km (supercritical) is the Moho reflection, which crosses the other body wave arrivals at about 170 km. At a range of 128 km a Rayleigh wave is seen to emerge at 24 seconds, traveling most strongly to the left but also to the right. The mechanism is identified as scattering of the Moho reflected P-wave by a 0.4 km step in topography (modeled by eight elements).

Figure 3 shows unclipped Rayleigh wave (R_g) arrivals across the model. The wave propagates about 2.65 km/s and travels halfway across the model when the calculation is terminated at 35 s.

The mesh is adequate to support 3–4 Hz fundamental mode Rayleigh waves so the observed dispersion of the 2 Hz wavelet (containing frequencies above 4 Hz) is partly due to numerical dispersion in addition to the more significant effects of lateral and vertical velocity variations near the surface. This figure clearly shows the relative strength of the Rayleigh wave compared to body waves for a surface pressure source, but hides all details. Of more use is the magnified response in Fig. 2. The middle synthetics show Rayleigh wave scattering by transitions in topography occurring between 12 to 25 km in range, and by lateral changes in wave speed at 33 and 55 km. Since the model's left side is assumed to be an axis of symmetry, apparent Rayleigh wave reflections are also seen. These are from the symmetric scattering features on the other side of the axis. This is of course a drawback of the axisymmetric approximation, however, the nonphysical reflected phases are easily identified. In general, this large-scale 2-D simulation produces a variety of propagation phenomena that would never be seen in a classical layered half-space analysis but is observed in real data.

5. Large-scale 3-D simulations of a scattering experiment

The second example calculation considered here models some rudimentary aspects of 3-D scattering experiments, performed by Reinke and Stump [6], in a small-scale alluvium site using five pound buried charges with seismometers and accelerometers placed around the shot out to 20–30 meters. One purpose of the experiments was to investigate azimuthal variability of ground motions by careful instrument calibration, exploration of site inhomogeneities, and the use of repeated axisymmetric shots. The principal question of interest here is the scattering produced by caliche lenses (cemented deposits) in the near-surface alluvial layer, and in particular, the interaction of characteristic lengths, i.e., wavelength, layer depth and size of the scattering feature. The following calculation was conducted to examine the feasibility of 3-D calculations to evaluate scattering by cemented lenses.

5.1. The discrete model

The geologic model for this case, illustrated in Fig. 4, is a $30 \times 30 \times 6$ meter quadrant of the explosive testbed, including two layers of alluvial material—a 3 meter low-speed surface layer over more competent alluvium. The finite element model is composed of $200 \times 200 \times 30$ finite elements, for a total of 1.2 million elements or 3.76 million nodal equations of motion. The model is truncated on the bottom by an absorbing boundary condition, and similarly on the outer boundaries, while the inner boundaries are symmetry planes. Elements in the lower layer are elongated in depth by a factor of two.

In real geologies, there are typically many scatterers with assorted shapes and orientations over the field of interest. For the purposes of this calculation, i.e., to determine propagation and source parameters for 3-D scattering simulations, a single ellipsoidal lens was modeled in the top layer, with nominal dimensions of $1 \times 3 \times 5$ m and oriented as in Fig. 4. The P-wave speed ratio between the caliche lens material and surrounding alluvium was taken as 2.6, based on field measurements. No attempt was made to smooth the lens boundaries using skewed elements, although comparisons of rough versus smooth scatterers should be made in future simulations.

The source in the field experiments is a five pound chemical explosive at the bottom of a two

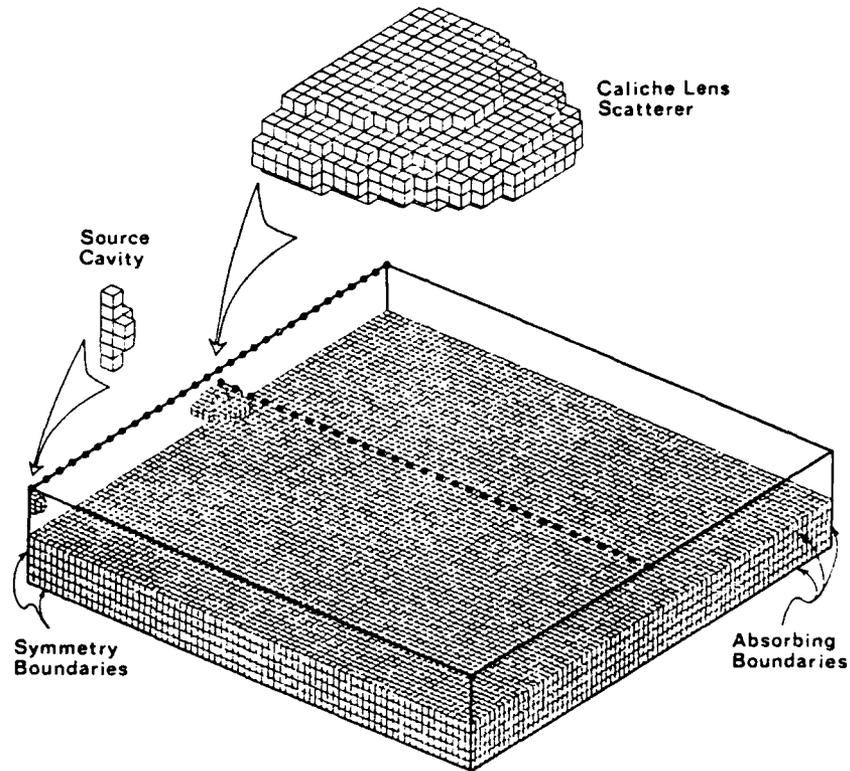


Fig. 4. The 3-D finite element model of a $30 \times 30 \times 6$ meter quadrant of the scattering experiment, from [6]. Details of the source cavity and scatterer show actual discretization. The upper layer is shown transparent, and the lower layer discretization is drawn with only 75 elements on a side although 200 are used in the actual model.

meter borehole. Rather than simulate this source directly, a simple pressure function was applied to a source cavity obtained by voiding elements around the symmetry axis in the model, as illustrated in Fig. 4. The source's pressure history was a step function, from which responses to other source time functions are obtained by convolution, in the same manner as described above for the 2-D model.

5.2. Finite element synthetic seismograms

Three-component velocity time histories were recorded on two lines over the model's free surface. One extended from the source epicenter over the lens to the absorbing boundary (i.e., on the line of symmetry), and the other was perpendicular to this line, extending from above the lens centroid to the boundary. Vertical velocity synthetics are plotted in Fig. 5 on the radial and crossing output lines, with the radial line data scaled by range and the cross line data merely magnified.

These records show weak body wave phases followed by fairly strong surface waves. Some limited observations can be made from these results, e.g., on the radial line the lens is seen to scatter surface waves into body waves downstream, but interpretation is complicated by the

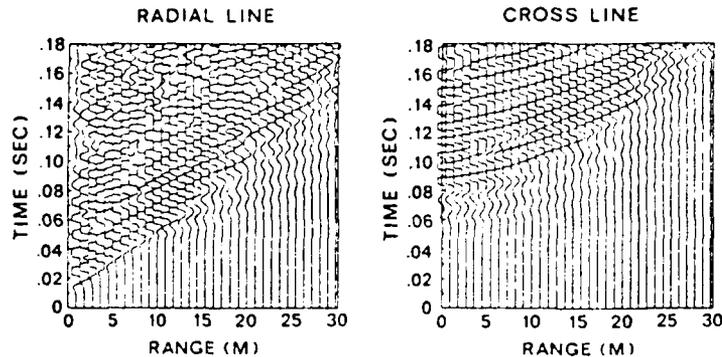


Fig. 5. Vertical velocity seismograms recorded on the cross and radial lines indicated by dots on the surface in Fig. 5. In-plane components of velocity were also recorded for polarization studies.

superposition of incident and scattered fields in the simulation. To overcome this it is necessary to run an additional free-field CRAY-2 simulation of the same geological model, i.e. with the scattering obstacle replaced by the surrounding medium. Subtracting seismograms from the two calculations yields the scattered field of interest. Of course this doubles the cost of analysis, however, for scattering studies it is essential for quantitative interpretation to decompose the incident and scattered wave fields.

Since the source geometry was rather coarse in this model, an additional free-field calculation would also be useful in evaluating azimuthal uniformity of the radiation pattern. In addition, a 2-D axisymmetric, free-field simulation could be used as a basis for comparison. Therefore, with the caveat that an additional free-field, 3-D simulation is required and a 2-D axisymmetric simulation is useful, these results indicate that this type of 3-D geophysical modeling is sufficient to produce useful data for experiment interpretation and planning. In order to draw further conclusions from the synthetic data in Fig. 5 it is necessary to perform further simulations. These are currently being done.

6. Seismic source modeling

One drawback of the models considered above is their inability to represent details of the source, since in large-scale simulations this region is only covered by a few elements. A solution is to either refine the mesh in the source region or perform a local source simulation and couple it to the global mesh. The latter approach is examined here, namely, the use of separate source simulations in highly refined models of the source region. Both linear and nonlinear models are considered.

Two types of source problems are modeled. The first includes 'nearly linear' surface or down-hole sources that apply vibratory or impulsive loads. These can be reasonably approximated by prescribing forces or velocities at an interface, with linear constitutive behavior assumed in the medium. The second type is the explosive source, for which strongly nonlinear material behavior dominates the near-source response. The explosive source involves a significant modeling effort to incorporate the proper energy release and coupling, in addition to difficulties associated with nonlinear constitutive behavior.

6.1. Elastic source modeling

A few examples of near-field seismic source models are examined here, assuming elastic properties. The idea is to model soil geometry around the source (mechanical vibrator, airgun, etc.) in some detail, and apply suitable pressures or velocities at the interface. The simplest geometry is the homogeneous half-space with prescribed surface pressure distribution and time function. A slightly more complex geometry is the borehole on the axis of an axisymmetric half-space, with either normal or shear tractions applied to all or part of the hole's boundary.

Only simple wavelet time functions are considered, representing the dynamic part of a total load. This total typically includes a significant static component, e.g., due to weight of the mechanical vibrator or sonde, which is ignored here since dynamic signals are of principal interest. It is permissible to have tension in the wavelet pressure boundary loading since this would be biased to net compression by the neglected static component.

The first example is an axisymmetric pressure loading over a small area of the free surface of a homogeneous 100×100 element model. The time function is a wavelet with center frequency chosen so that about 20 elements support it. A Gaussian spatial distribution of lumped vertical forces is assumed near the model axis. A sequence of vector snapshots shows the resulting velocity wave field in Fig. 6. The left side is the axis of symmetry, the right and bottom sides are absorbing boundaries, and the top is the free surface. The Gaussian pressure contributes significant forces only on the leftmost three or four nodes. Note that the simple absorbing boundary is quite effective for the P-wave, but less so for the S-wave at shallower angles, as well as for the Rayleigh wave.

To quantify this refined source solution or incorporate it in larger models, a useful approach is to record velocity time histories on a surface surrounding the source in its far field. For convenience this is typically a Cartesian surface (rectangle or box) at 10–20 characteristic source dimensions. In Fig. 7, results from the present calculation show velocities at a takeoff angle increment of 15° . These are rotated to display radial and tangential motion. This approach yields a simple spatial characterization of the source function. By calculating the impulse or step response instead of a particular wavelet, other time functions can be obtained by convolution.

The remaining elastic examples considered here involve pressure or shear loading on a section of a borehole. Figure 8 shows a snapshot sequence for a 75 ft hole with a pressure wavelet applied over a 6 ft segment centered at a depth of 37 ft. Similarly, Fig. 9 shows a sequence for the same geometry with a shear wavelet applied over the segment. The loads in both examples transition from zero to full traction over one element, rather than tapering. This is too strong a discontinuity for a discrete model to comfortably support, and consequently some spurious local oscillations (so-called hourglassing) are excited. Hourglassing is automatically removed at the element level by using orthogonality of the element's model shapes, e.g., see [7].

The above results are useful in comparing different source types or in characterizing non-ideal sources for input to a ray tracing model. These data are also necessary in order to interpolate input to coarser models.

6.2. Inelastic source modeling

There are various approaches to simulating underground chemical explosions. One is to apply a pressure history to the boundary of a cavity defined at some suitable 'elastic' radius; another is

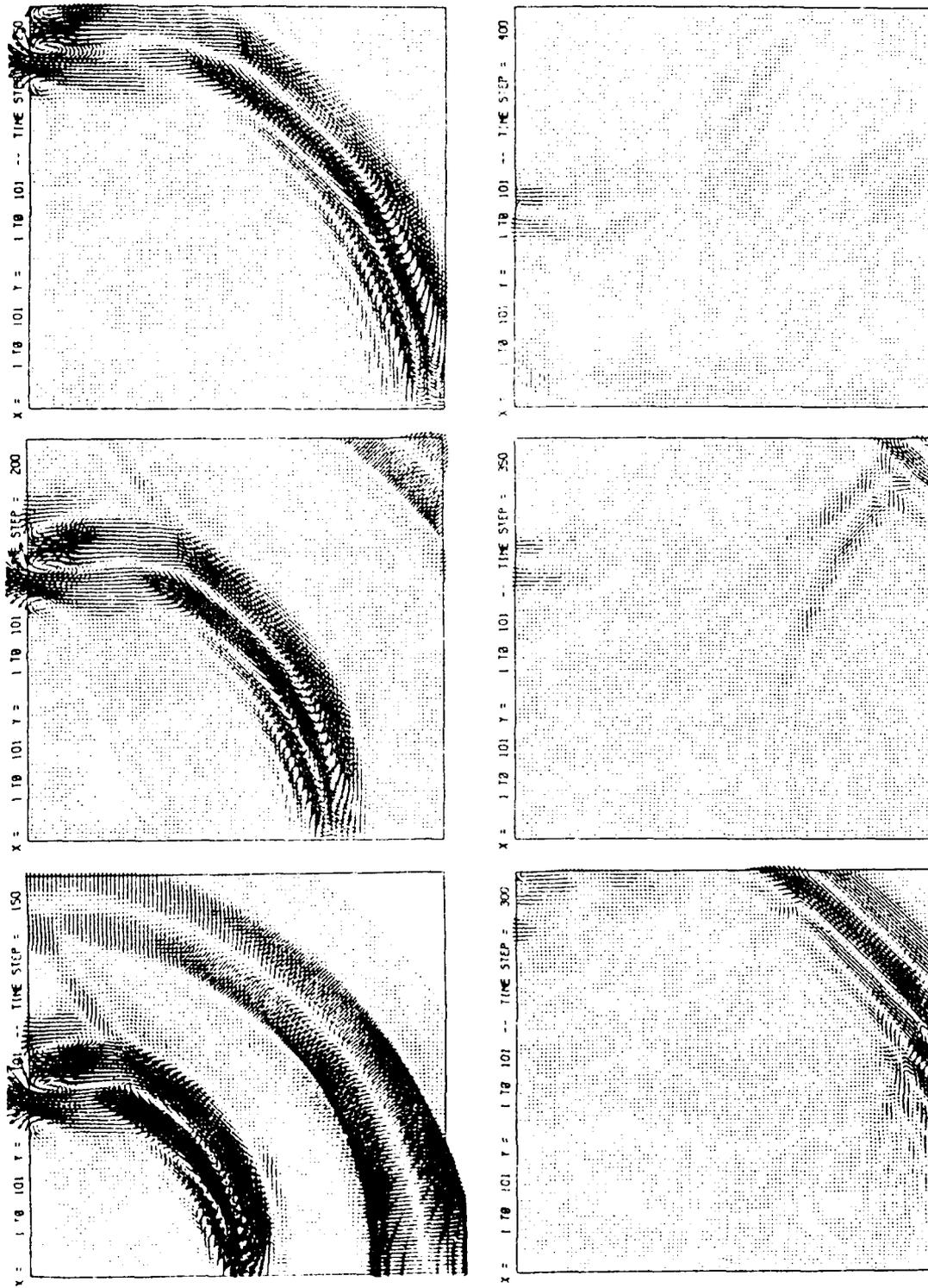


Fig. 6. A sequence of velocity vector snapshots showing P_s , SV-, and Rayleigh waves emanating from a Gaussian surface pressure (on the top) centered on the axis of symmetry (left side). The right and bottom boundaries have radiation conditions. The sequence starts at 150 time steps with a 50 step increment.

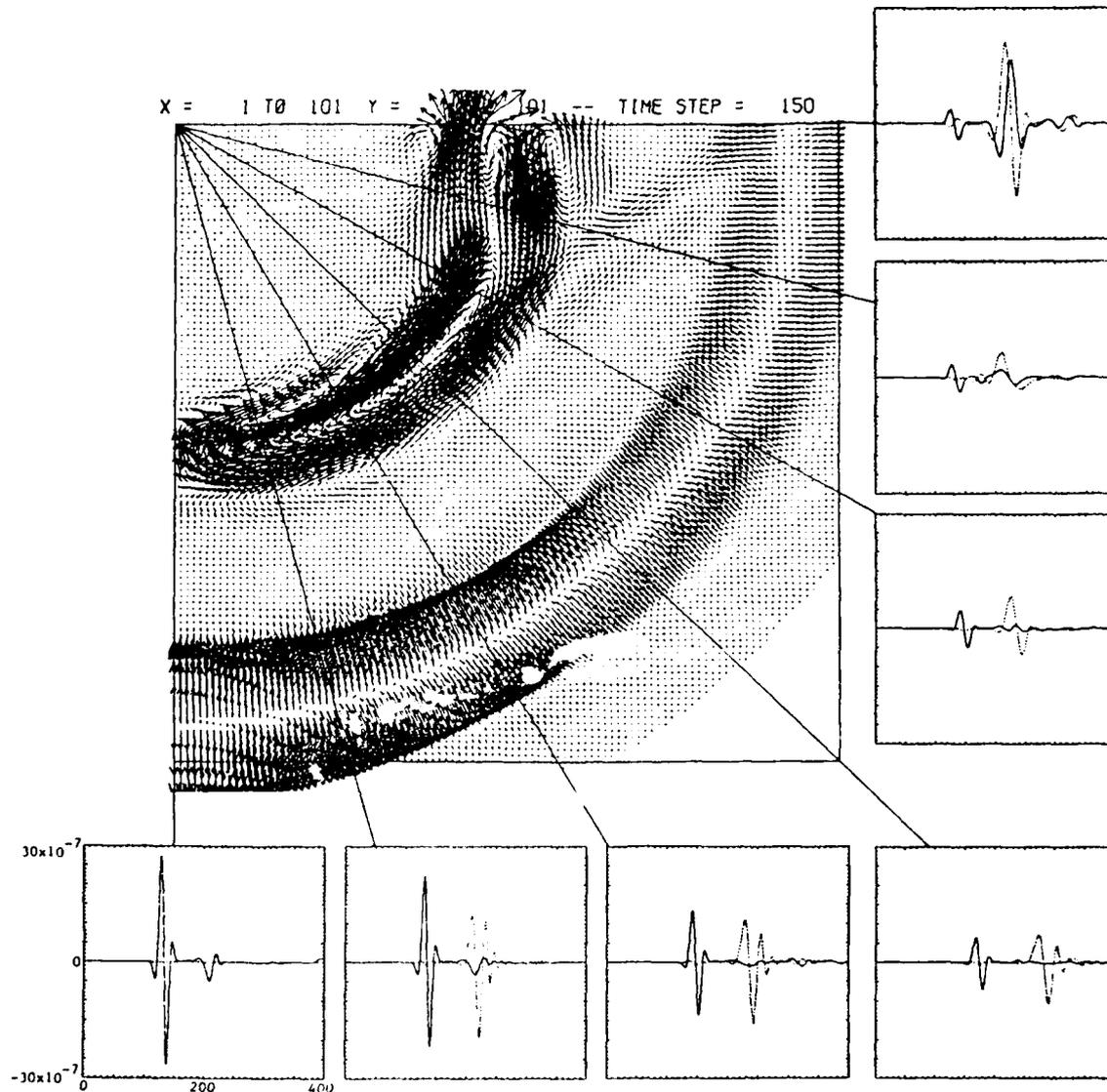


Fig. 7. Velocity time histories versus takeoff angle for the calculation in Fig. 6. Amplitude is normalized to a unit radius on radial lines from the center of the source region at 15° increments. The solid and dashed curves show radial and tangential motion, respectively, and quantify the source's radiation pattern.

to apply suitable initial conditions on velocity and pressure over an effective source region; and still another is to model the expanding gases from the source starting from the initial explosive volume. Of course the level of modelling detail can be further increased to include explosive dynamics, vaporization, etc. More rigorous modeling is seldom necessary unless local effects like cavity growth or cratering are important.

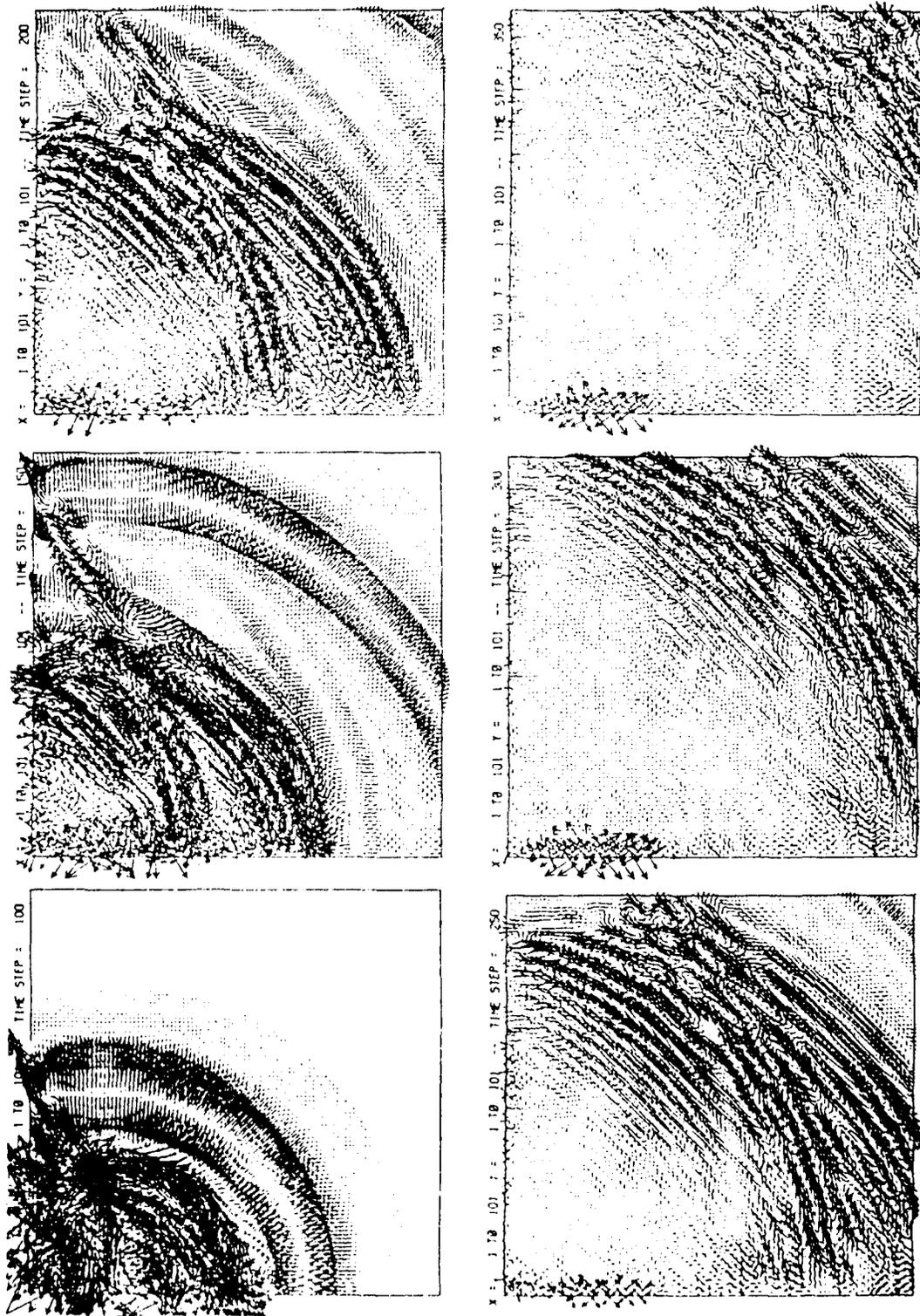


Fig. 8. A vector velocity snapshot sequence for a pressure wavelet applied to the sides of a 75 ft borehole at its mid-depth. The model is 200 x 200 ft. The pressure discontinuity causes significant spurious oscillation on the cavity boundary.

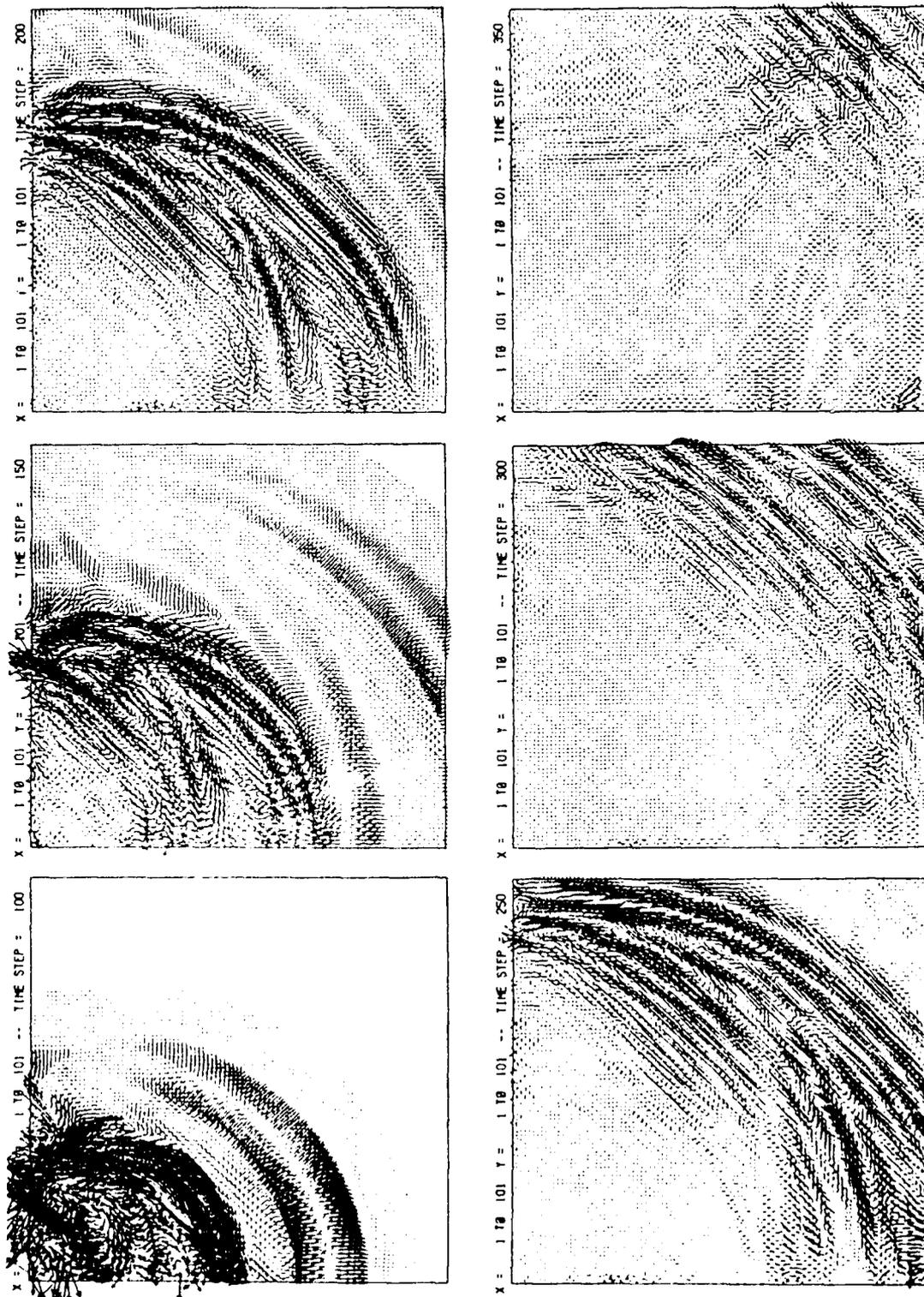


Fig. 9. A vector velocity snapshot sequence for a shear wavelet applied to the sides of a 75 ft borehole at its mid-depth. The model is the same as that for Fig. 8. The shear discontinuity is seen to cause much less spurious oscillation than the pressure discontinuity.

If no information is available on nonlinear constitutive behavior of the near-source medium, then the only alternative is to prescribe pressure or velocity on an 'elastic' cavity boundary. Details of the time function are either chosen to populate an observed frequency spectrum, or determined from experimental data in similar media. If a nonlinear model is available then bulk initial values can be used to incorporate additional source physics in the simulation. The nonlinear basis for the present source simulation is the cap model.

Chemical explosives typically exhibit confined detonation pressures of 2.9×10^6 psi (200 kbars) or less. Cap parameters are defined for pressures up to 200,000 psi in the alluvial-type media of interest here. Therefore, the cap model cannot represent material behavior in the immediate neighborhood of the detonated explosive. It is necessary to determine a radius where pressure has decayed to 200,000 psi and prescribe initial values over this volume determined from shock jump conditions. This is the basis for the bulk initial value approach used here. It has

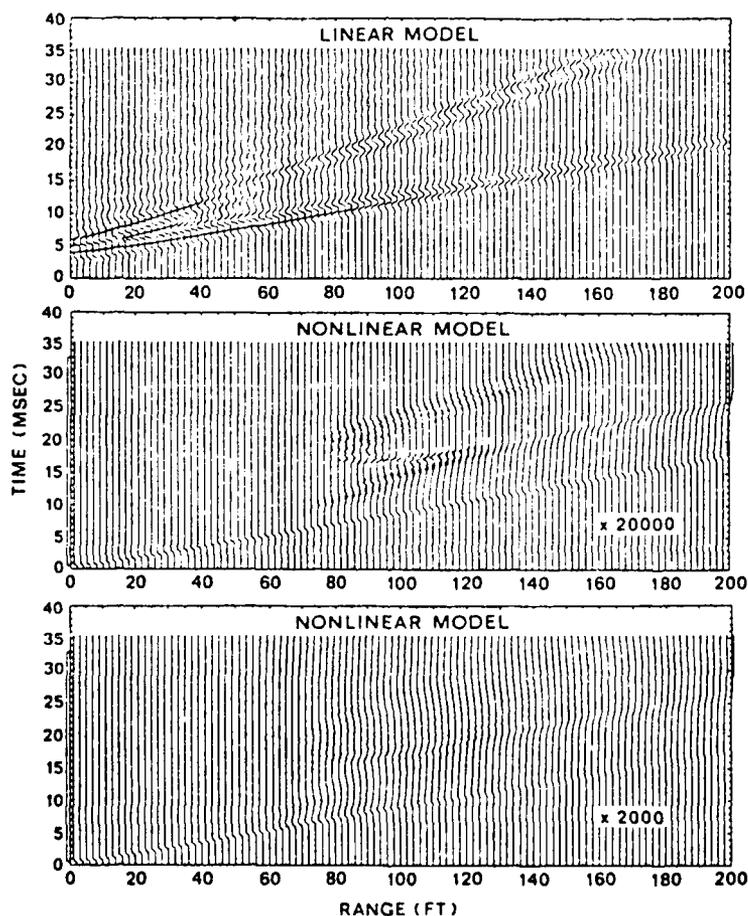


Fig. 10. Suites of vertical velocity seismograms for the nonlinear and linear explosive source simulations. Upward velocity is plotted to the left. For ranges less than 85 ft the response is dominated by surface spall. Beyond this range nearly linear response is observed, yielding 30 Hz single cycle wavelet-type motion.

been used extensively by Sandler [8] and validated against a variety of explosive source experiments.

The model examined here consists of a vertical cylinder of explosive 10 ft long and 0.25 ft in diameter, with the upper end buried at a depth of 10 ft in homogeneous alluvium. Assuming a nonlinear pressure decay proportional to $1/r^2$ (for cylindrical geometry) yields the 200,000 psi radius at about 0.5 ft. Therefore, the volume over which initial conditions are prescribed is 1 ft in diameter and 11 ft long. This volume of material is initially pressurized to $P_0 = 200,000$ psi, with initial radial velocity set to zero on the axis and increasing linearly to V_0 at $r = 0.5$ ft, where V_0 is the particle velocity determined from the jump condition for a shock pressure of P_0 and a shock speed based on the secant modulus.

The finite element model examined here covers a 150 ft deep by 200 ft wide axisymmetric region of homogeneous material (alluvium) discretized into 151×200 elements. On the model axis, elements are 0.5 ft wide so that initial conditions are prescribed over a single column eleven elements long. Figure 10 shows vertical velocity synthetic seismograms for both nonlinear and linear simulations. The two lower suites are magnifications of the nonlinear calculation, while the upper shows elastic response for a low-pressure wavelet applied to the cavity boundary.

Since the explosive source is very energetic, soil immediately above and to the side of the model axis is spalled, i.e., launched at high velocity with complete tension cutoff. This clearly violates the assumptions of small strain theory and no claim is made that this late time behavior is realistic. However, the waves of interest have propagated beyond the source region before mesh distortions and constitutive uncertainties invalidate the solution. In any event, at a range of 85–90 ft, vertical velocity has decayed to a small fraction of that over the shot. This behavior is shown in the lower suite and magnified by ten in the middle suite to show additional details. Upward velocity is drawn to the left in the individual seismograms. Referring to the middle suite, beyond 110 ft the sequence of phases are, first a P-wave with upward motion (loading) arriving at the elastic wave speed, then a stronger P-wave about 10 ms later with downward motion (unloading), and followed by an S-wave with upward motion again. This up-down-up motion is suggestive of the common velocity wavelet approximation used as source time function in geophysics. Integrated yields a 30 Hz up-down (sinusoidal) cycle in displacement for the present nonlinear case.

The elastic simulation shown in the upper suite confirms that the slope of the phases in the nonlinear synthetics correspond to elastic wave speeds of the medium. Center frequency of the wavelet is 300 Hz, which is nonphysical but chosen to maximize separation between the phases while minimizing grid dispersion. Clearly, the frequency is not high enough to separate the shear and surface wave since they differ by about 8% in phase velocity for such geologic media.

7. Discussion and conclusions

This paper has described a number of discrete wave modeling problems in geophysics, including the simulation of large-scale 2-D refraction experiments, small-scale 3-D scattering experiments, and axisymmetric seismic sources, both linear and nonlinear. All of these problems require numerical solution because some level of nonseparability or nonlinearity precludes the use of classical analysis techniques.

Since the finite element algorithms employed operate at the lowest level of numerical sophistication practical for the problems at hand, i.e., linear, Cartesian elements and explicit

(leapfrog) integration, very little attention has been paid to theoretical background here since it is well documented in the finite element and finite difference literatures. One exception is the cap model for implementing nonlinear constitutive behavior, and so some theoretical background is provided on this topic. The principal aims of the paper are to demonstrate the simulation and performance capabilities of the CRAY-2, possibly the fastest and largest general purpose machine available today, and to familiarize the theoretical analysis community with some of the practical modeling issues and problems encountered in geophysics. These issues include (1) the capacity for large-scale inhomogeneous models and their efficient implementation, (2) very fast processing with very large, fast memory, and (3) capability for nonlinear behavior, material attenuation, and radiation boundary conditions.

7.1. Large-scale modeling capacity

Regarding CRAY-2 machine capacity for large-scale simulation, we can comfortably assert that 2-D models are well-resolved and appear cost-effective for most practical linear propagation problems, however, 3-D models remain resource limited and are not cost-effective. Here the term cost-effective means that simulations cost much less than the physical experiments. The reason for the disparity between 2-D and 3-D capability is that, for explicit calculations, computational and memory requirements grow like n^{d+1} where n is the number of nodes in a representative direction, and $d + 1$ includes the problem's spatial dimension (d) and time dimension. For example, if it is necessary to double a model's size or mesh resolution (halve the spacing), then resource requirements increase by a factor of 8 in 2-D and a factor of 16 in 3-D. This additional factor in 3-D is all too often prohibitive, not necessarily in terms of memory but more often in terms of execution time.

In terms of the 3-D scattering model described above, the frequency resolution (element size) is adequate but the symmetry conditions and depth are too restrictive to reproduce the full experiment. As a consequence the model's size needs to be effectively doubled. For this more realistic model the execution time would increase from 1.5 hours to 24 hours, while memory requirements would increase from 20 million words to about 160 million. For this grid a numerical simulation would probably cost twice as much as the experiment.

In the case of the 2-D refraction model considered above, the original discretization had twice the element size of the final mesh. Unfortunately, this coarse model would not support wavelet frequencies much above 1 Hz, and since the experiments indicated at least 4 Hz signal resolution the element size was halved as a compromise. Consequently, the execution time went from an original estimate of 1 hour to over 8 hours, and memory quadrupled from 5 million to 20 million words. However, this simulation still cost a small fraction of one of the original refraction experiments.

7.2. Speed and memory

The principal requirement for the large-scale simulations described in this paper is minimizing disk I/O by retaining the entire model in fast memory. Thus, very large memories are necessary and the 256 million words available on the CRAY-2 are adequate. However, there is a hardware mismatch between scalar fetch and vector processing speed, which can seriously degrade performance. Fortunately it can be circumvented by using less sophisticated inhomogeneous

model representations at the expense of memory. This puts a serious limitation on problem size when the memory limit is 2 megawords (early CRAY-1), but is insignificant with 256 megawords available. Experiments with very large problems requiring upwards of 200 megawords show that there is no degradation in performance as memory usage approached the limits of the machine.

The major issue in large-scale, explicit wave simulation is processing speed. In order to do practical 3-D modeling, at least an order of magnitude increase in speed is necessary, for the reasons cited above. Furthermore, this capability should not cost much more than present CPU charges (\$1000 to \$3000 per hour say), otherwise we could use the CRAY-2 as it stands and pay \$25,000 to \$50,000 per calculation. The point is to make 3-D calculations timely and affordable by increasing speed while holding cost constant. The digital computer's forty year history and present state shows that this is not an unreasonable expectation.

There are two avenues available to increase processing speed: one is to increase the central processor's clock rate and the other is to increase the number of processors. A factor of three increase in speed (faster clock, more efficient vector processors) can be reasonably expected in the next generation of CRAY machines, and by using the four available processors a factor of 10 could probably be achieved. Unfortunately, the cost would be prohibitive since these multiprocessor supercomputers are in reality multicomputers, essentially four computers in one, with the cost directly proportional to the number of processors used. The alternative is to use a dedicated multiprocessor machine with slower processors but many more of them. This solution of course depends on the ability to logically partition the problem, assign pieces to the processors, avoid memory conflicts, etc. These are currently research questions but will be addressed soon since machines are available and there exist large-scale problems demanding solutions. However, it is not clear that CRAY-type machines will provide the necessary hardware and software.

7.3. Nonlinear behavior, material attenuation, and boundary conditions

We gather here some comments on other technical issues illustrated by the source models described above. These linear and nonlinear models are local refinements of the source regions in large-scale simulations, and all are in homogeneous media. They were readily calculated on a minicomputer rather than a CRAY.

The ability to model nonlinear material behavior is critical near energetic sources such as explosions. This is not to say that mechanical devices do not involve some level of nonlinearity, only that it does not obviously dominate the solution. Note that the nonlinear behavior associated with medium to low stress levels has not been characterized to the extent that high dynamic stress levels have. The important feature of nonlinear modeling with the cap and similar plasticity algorithms is their extensive use of stress state tests and logical branching. This branching effectively prevents vectorization of the algorithm. Consequently, nonlinear geophysical simulations are not vectorizable on CRAY-type supercomputers, hence are quite slow. An alternative is to use parallelism to speed up the cap algorithm, but this is not yet feasible on CRAYs.

None of the models considered here include anelastic attenuation, i.e., frictional damping, in the material behavior. Although this feature is straightforward in implicit frequency domain simulations, it is difficult to incorporate in the time domain. An implementation is available in FLEX, using a standard linear solid algorithm tuned to give constant damping over a given range

of frequencies. However, the requisite memory and arithmetic operations exact a significant cost. Such ad hoc attenuation models should be used with caution when signal attenuation is an important feature of a time-domain simulation.

Regarding boundary conditions, it is clear that discrete wave simulation models must be truncated in space by suitable radiation boundary conditions. For the present examples, the lowest-order condition assuming normal incidence of plane P- and S-waves was generally adequate. Higher-order conditions valid for wider incidence angles are available [9], but in many cases are unnecessary since boundary reflections from the model's side can usually be identified in the synthetic seismograms.

Since virtually all theoretical work on time-domain absorbing boundaries assumes a homogeneous medium, any implementation is degraded when properties vary along the boundary, e.g., on the side. This depth variation is the rule in geophysics and should be accommodated in future theoretical work. Some care must also be exercised on the model's lower boundary because the typical increase of propagation speed with depth in geologic models causes waves to turn away from the bottom, becoming more grazing there and less efficiently absorbed. Finally, note that radiation conditions in nonlinear models present a host of new difficulties.

7.4. Conclusions

The calculations described here give an indication of our present ability to simulate large-scale, time-domain wave propagation in nonseparable or nonlinear models. They show that one of the most powerful supercomputers available today, the CRAY-2, can indeed perform very large, 2-D, elastic simulations efficiently, despite an imbalance between scalar memory access and vector processing speeds. Of course this assumes that the principal processing loops are fully vectorized. Explicit, linear wave solvers are well suited to vectorization, in contrast to nonlinear constitutive algorithms for example.

The calculations also indicate that, for 3-D problems commonly encountered in geophysics, still more speed is required. This is not to say that we cannot do useful 3-D simulations with the CRAY-2. In fact the above scattering example shows that some limited but interesting 3-D simulations in geophysics can finally be addressed at reasonable cost. What must be emphasized is that cost grows so quickly in the third dimension, that only processor speed increases by one, or better yet, two orders of magnitude, will push affordable model size beyond the minimum required for adequate temporal and spatial resolution. Multiprocessor (not multicomputer) hardware with distributed algorithms (e.g., Connection type machines), or vector-parallel machines with suitable compilers (e.g., Alliant type), provide another avenue to achieving the necessary speed increase.

Simulations of the type described here are the only means of investigating propagation phenomena involving both range and depth dependent (nonseparable) or nonlinear model properties. Other options like ray tracing or boundary integrals involve idealizations that are unacceptable for this broader class of earth models. Note that discrete geophysical simulation has, here and elsewhere, been used principally for the forward problem. As the above performance issues are addressed more fully by software and hardware, it will eventually become possible to implement discrete models for the inverse problem, i.e., in an optimization loop with the synthetic data generated from discrete trial models. This possibility is perhaps the ultimate goal of geophysical simulation research.

Acknowledgment

We gratefully acknowledge support for this research by the Air Force Geophysics Laboratory under contract F19628-84-C-0102, and computer time on the CRAY-2 generously provided by the University of Minnesota's Supercomputer Institute.

References

- [1] D.K. Vaughan, FLEX users guide. UG8298, Weidlinger Associates, Palo Alto, CA, 1983.
- [2] K.-J. Bathe and E.L. Wilson, *Numerical Methods in Finite Element Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1976).
- [3] I.S. Sandler and D. Rubin, An algorithm and a modular subroutine for the cap model, *Internat. J. Numer. Anal. Meth. Geomech.* 3 (1979) 173-186.
- [4] S.L. Klemperer and J.H. Luetgert, A comparison of reflection and refraction processing and interpretation methods applied to conventional refraction data from coastal Maine, *Bull. Seism. Soc. Amer.* 77 (1987) 614-630.
- [5] J.H. Luetgert, Personal communication, United States Geological Survey, Menlo Park, CA, 1987.
- [6] R.E. Reinke and B.W. Stump, Stochastic geologic effects on near-field ground motions in alluvium, Submitted for publication.
- [7] D.P. Flanagan and T. Belytschko, A uniform strain hexahedron and quadrilateral with orthogonal hourglass control, *Internat. J. Numer. Meth. Engrg.* 17 (1981) 679-706.
- [8] I.S. Sandler, Personal communication, Weidlinger Associates, New York, 1987.
- [9] B. Engquist and A. Majda, Radiation boundary conditions for acoustic and elastic wave calculations, *Comm. Pure Appl. Math.* 32 (1979) 312-358.

A MATHEMATICAL THEORY OF LIGHT DIFFRACTION BY DIELECTRIC WEDGES[†]

by
Gregory L. Wojcik
Weidlinger Associates*

ABSTRACT

The diffraction of light by dielectric wedges occurs whenever electromagnetic waves illuminate a sharp edge on an interface separating different dielectric media, e.g., air and glass. Such edges are often encountered in practice, for example, on prismatic and faceted transparent objects, or, more generally, on nonconducting cylindrical and conical structures with polygonal cross-sections. However, despite its common occurrence and intrinsic importance, dielectric wedge diffraction has never been quantified due to the lack of an effective mathematical theory. The reason is that, since distinct waves are coupled across the edge, their diffraction is fundamentally a nonseparable, vector field phenomenon, hence, very awkward mathematically.

This paper presents a new vector theory for the planar canonical problem of two, contiguous, semi-infinite, dielectric wedges, with a TM or TE polarized plane wave incident from the larger wedge and parallel to the edge. The mathematical formulation, based on self-similar solutions of the vector wave equation in two space dimensions, yields a vector initial-boundary problem in connected complex half-planes. Solutions are found in terms of Cauchy-type integral representations on the boundaries. Boundary conditions yield a singular integral equation to be solved for the kernel of the boundary integral representation. The theory is developed for an example that minimizes complexity while exhibiting all of the vector character, and evaluation of a verification problem is discussed.

[†] Supported under AFGL Contract F19628-84-C-0102 and NSF SBIR Grant SI-8760089

* 4410 El Camino Real, Suite 110, Los Altos, CA 94022

DESCRIPTION OF THE PROBLEM

1. Introduction

The phenomenon of dielectric wedge diffraction is encountered whenever electromagnetic radiation illuminates a sharp edge on an interface separating two (or more) distinct dielectric media. Many physical problems exhibit this type of diffraction, for example, the scattering of millimeter (radar) waves from nonconducting cylindrical or conical objects with polygonal cross-sections, or less conventionally, the scattering of micrometer (optical) waves from dielectric prisms or faceted structures on semiconductor substrates. The phenomenon is particularly relevant to imaging by means of diffracted radiation, e.g., radar cross-sections of nonmetallic, angular targets, and optical metrology of small dielectric features.

Yet, despite the importance of dielectric wedge diffraction in many optical and radar applications, no mathematical theory currently exists, either in physical optics or applied mathematics, capable of solving the problem. The reason is, since waves with different speeds are coupled across the interface at the edge, this is a vector rather than scalar wave phenomenon and none of the mathematical formulations attempted over the last century are able to accommodate the singular coupling of two or more, i.e., vector, wave functions on an interface with discontinuous curvature. Consequently, there has been no effective way to satisfy boundary conditions globally. Mathematical intractability persists even in the immediate neighborhood of the edge, when global curvatures or other "characteristic length" interactions can be ignored.

In what follows, the canonical initial-boundary problem governing plane dielectric wedge diffraction is formulated and solved under the assumptions of classical field theory. This treatment is appropriate provided the incident radiation is adequately represented by waves rather than quanta. The mathematical problem is one of generalized vector diffraction since it considers a generalized interface, i.e., an infinitely sharp edge on a plane interface, separating two different media supporting wave fields. The generalized two-wedge model consists of exterior and interior, two-dimensional, semi-infinite domains, with the larger wedge defining the exterior. It is assumed that a plane, polarized electromagnetic wave is incident on the edge and parallel to it, with either the electric component (TM polarization) or the magnetic component (TE polarization) directed along the edge. For simplicity it is also assumed that the wave is incident from the larger exterior wedge and does not reflect from the interface until it touches the edge. Clearly, since there is no spatial or temporal variation in the geometry or incident field along any line parallel to the edge, the problem's formulation may be confined to any perpendicular plane.

Some examples of generalized vector diffraction are illustrated in Fig. 1, showing incident, reflected, transmitted, and diffracted wave fronts. The four cases drawn are combinations of two interior wedge angles, 120° and 60° , and two relative wave speeds, $c_1 < c_2$ and $c_1 > c_2$, where c_1

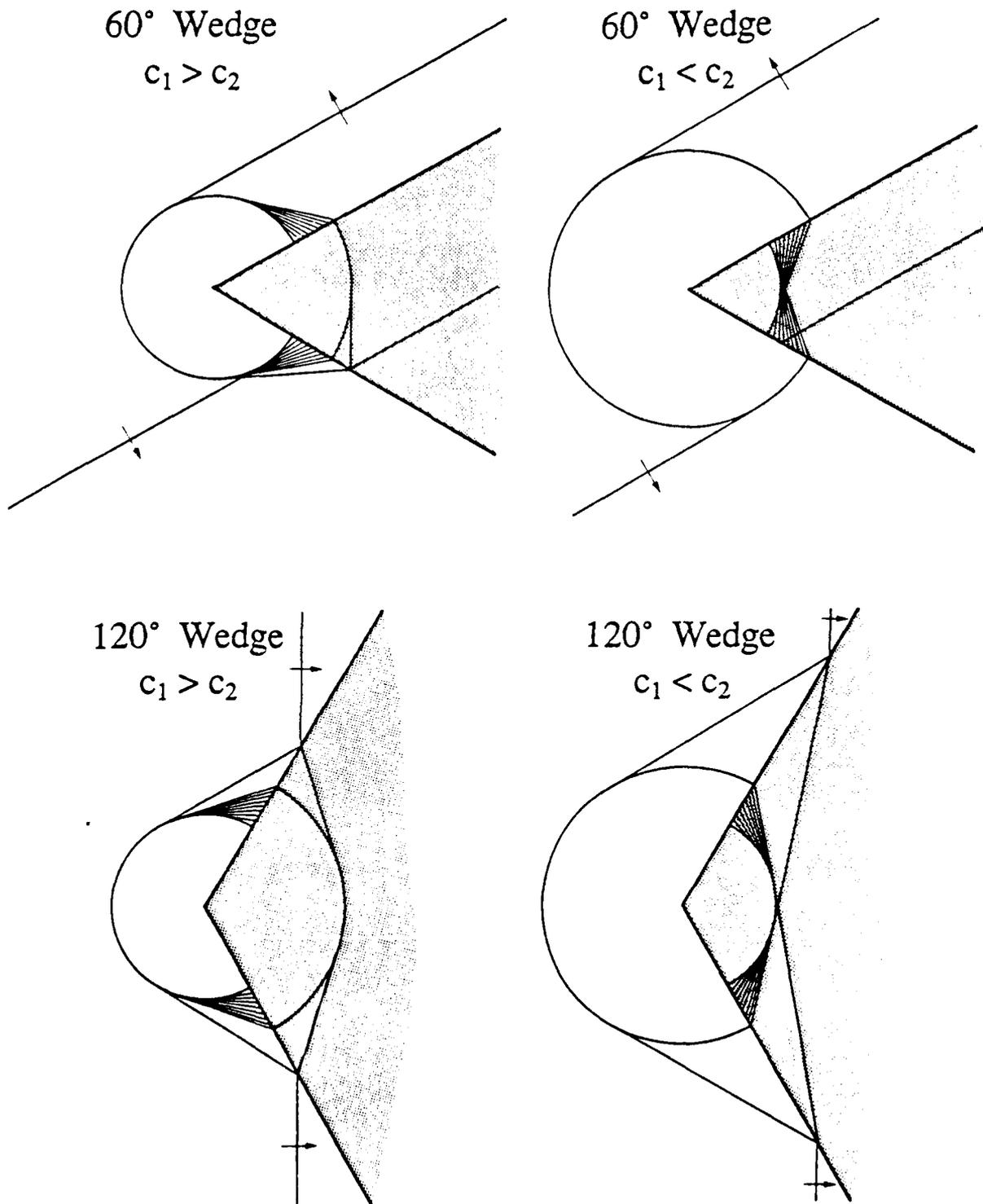


Figure 1. Examples of dielectric wedge diffraction for plane wave incident on 60° and 120° wedges.

and c_2 are speeds in the interior and exterior wedges, respectively. To illustrate some limiting examples, the incident wave fronts are drawn nearly perpendicular to the line of symmetry in the 120° wedges, and parallel to the upper interface in the 60° wedges. Note, as a general rule all plane wave fronts propagating at c_1 or c_2 in the interior and exterior wedges, respectively, are tangent to their supporting wedge's diffracted wave front or its "virtual" extension into the other wedge. Thus, it is relatively simple to draw wave fronts for any incidence angle by continuing tangentially off of the diffracted circles. Referring to the figure, the circular wave fronts constitute the leading diffracted wave and the "triangular" wave fans are Mach waves excited in the slower wedge by the faster diffracted wave grazing the interface. For the case of $c_1 < c_2$, as the interior wedge angle decreases, the Mach wave fans will intersect, eventually reflecting between the interfaces. This yields the most complicated situation in terms of analysis. The case of $c_1 > c_2$ is generally simpler, although not as relevant to optical applications. In order to simplify the presentation, this simpler case will provide the model for much of the theory's development.

The canonical vector diffraction problem illustrated in Fig. 1 is well-known in applied mathematics, but has remained unsolved since its recognition in electromagnetism and elasticity perhaps a century ago. In contrast, the associated scalar diffraction problems have been completely solved. In 1896 Sommerfeld [8] gave a solution in the degenerate scalar wedge (2π included angle), better known as the opaque screen. MacDonald [4] formally solved the first nondegenerate scalar problem in 1902, with the first uniform asymptotic expansion, i.e., valid everywhere in the domain, obtained by Pauli [6] in 1938. A variety of more direct scalar solutions have since been derived, based primarily on separation of variables or transform methods, but also on the principle of self-similarity, e.g., Keller and Blank [3]. In contrast to scalar diffraction, vector diffraction involves two or more waves in one or more wedges, and the essential boundary coupling in wedge-shaped domains defies separation of variables. This leaves the principle of self-similarity, or equivalently, the theory of homogeneous solutions, as the most viable mathematical approach.

A study [10] by the present author has finally derived an effective self-similar vector theory. There are two essential steps. First is transformation of the vector wave equation in the generalized diffractor via self-similarity. This yields d'Alembert's classic string equation, i.e., the 1-D wave equation, in characteristic coordinates, which is solved by inspection. The solution is mixed, that is, hyperbolic or elliptic over different parts of the domain, so substitution of the boundary conditions yields a mixed boundary problem in semi-infinite strips. The second essential step is transformation of the mixed strips to two complex half-planes by the so-called characteristic mapping, essentially a Schwarz-Christoffel transformation, under which boundary conditions and the real characteristics are mapped entirely to the real axes. By eliminating the hyperbolic solutions, nonmixed elliptic boundary conditions are prescribed. This reduces the boundary problem to a vector form of the Riemann-Hilbert problem on the real axes of two complex planes.

By means of Cauchy-type integral representations, the Riemann-Hilbert problem reduces to a singular integral equation that must be solved on the two real lines.

The mathematical approach developed here is elegant, and straightforward once certain key features are appreciated. It provides a consistent, constructive procedure for solving this most difficult diffraction phenomenon. Of course, the problem's long history of intractability proves that no formal analysis should be taken for granted, hence the need for numerical solution of the integral equation to ultimately confirm the theory. Therefore, to validate the solution, numerical results for the particular case of the canonical problem with $c_1 > c_2$ are discussed briefly at the end of the paper. Although an all-encompassing analysis, valid for every combination of geometry, input, and wave speed is not presented, the techniques developed here are necessary and sufficient to analyze all such situations. However, these details must be left to future papers not burdened by detailed development of the underlying mathematical theory.

The paper is presented below in twelve sections organized under five logical headings. Under DESCRIPTION OF THE PROBLEM, an introduction is provided (Sec. 1), Maxwell's equations are reviewed (Sec. 2), and the mathematical problem is stated (Sec. 3). Under WAVE SOLUTIONS AND THE INITIAL-BOUNDARY PROBLEM, self-similarity is introduced and homogeneous solutions are determined (Sec. 4), the resulting vector initial-boundary problem in semi-infinite strips is formulated (Sec. 5), and the characteristic mapping of the strip domains to two complex half-planes is applied (Sec. 6). Under REDUCTION TO ELLIPTIC BOUNDARY CONDITIONS, the hyperbolic part of solutions are determined (Sec. 7), and elliptic boundary conditions on the real axes of the half-planes are derived in terms of vector Riemann-Hilbert (R-H) problems (Sec. 8). Under INTEGRAL REPRESENTATIONS AND SINGULAR EQUATIONS, the scalar R-H problems are solved for the basic structure of global solutions (Sec. 9), the vector R-H problems are satisfied by Cauchy-type integral representations yielding a singular integral equation (Sec. 10), and some undetermined functions are evaluated (Sec. 11). Under DISCUSSION AND CONCLUSIONS, generalizations are discussed and a numerical solution of the canonical problem is described (Sec. 12), and conclusions are drawn (Sec. 13).

2. Maxwell's Equations

Electromagnetic waves in linear, isotropic, time-invariant media are governed by a simplified form of Maxwell's equations, namely

$$(2.1) \quad \nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \quad , \quad \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}$$

where \mathbf{E} and \mathbf{H} are the electric and magnetic field intensity vectors, respectively, while ϵ is the dielectric permittivity and μ is the magnetic permeability. This form assumes that the permittivity and permeability are constant, so that gradients can be neglected. Note that electric current and

charge terms have been omitted from (2.1) since they are not immediately relevant to the optical applications of interest. For the nonmagnetic media considered here, μ is essentially equal to its vacuum value, μ_0 , everywhere.

In two-dimensional, plane geometries it is convenient to decompose the electromagnetic field into transverse magnetic (TM) and transverse electric (TE) components, e.g., see [2], from which general polarizations may be obtained by superposition. In Cartesian coordinates, the TM polarization unknowns are E_z , H_x , and H_y , governed by Maxwell's equations (2.1) in the form

$$(2.2) \quad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \epsilon \frac{\partial E_z}{\partial t} \quad , \quad \frac{\partial E_z}{\partial y} = -\mu \frac{\partial H_x}{\partial t} \quad , \quad \frac{\partial E_z}{\partial x} = \mu \frac{\partial H_y}{\partial t}$$

and for TE polarization the unknowns are H_z , E_x , and E_y , governed by

$$(2.3) \quad \frac{\partial H_z}{\partial y} = \epsilon \frac{\partial E_x}{\partial t} \quad , \quad \frac{\partial H_z}{\partial x} = -\epsilon \frac{\partial E_y}{\partial t} \quad , \quad \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\mu \frac{\partial H_z}{\partial t}$$

In either case, taking appropriate time and space derivatives and eliminating the transverse field components yields the scalar wave equation

$$(2.4) \quad v^2 \Delta f = \frac{\partial^2 f}{\partial t^2}$$

where f stands for E_z or H_z , $v^2 = 1/\epsilon\mu$ is the square of the speed of light in the medium, and Δ is the Laplacian. Thus, the two polarizations are each governed by the scalar wave equation.

The above equations are valid provided there is no spatial variation in the permittivity. When ϵ is discontinuous across a material interface, certain continuity conditions must be applied there. These are derived by, typically, integrating Maxwell's equations over the surface of an infinitesimally thin "pillbox" surrounding part of the interface and applying Stokes integral theorem, e.g., [9]. Evaluating the integrals shows that tangential components of the electric and magnetic fields are continuous across the interface. For example, at a discontinuity in permittivity across the x - z plane, this means that E_z and H_x (or $\partial H_x/\partial t$) are continuous in the TM case, and likewise for H_z and E_x (or $\partial E_x/\partial t$) in the TE case. From (2.2), it follows that TM or TE boundary conditions on wave equation (2.4) are, in general,

$$(2.5) \quad f_1 = f_2 \quad , \quad v_1 \frac{\partial f_1}{\partial n} = v_2 \frac{\partial f_2}{\partial n}$$

Subscripts denote the medium and coordinate n is along a vector perpendicular to the interface into one medium. For the TM case, $f = E_z$ and $v_i = 1/\mu_0$, while for TE, $f = H_z$ and $v_i = 1/\epsilon_i$.

3. Problem Statement

The generalized, plane diffractor consists of two semi-infinite, contiguous dielectric wedges defined by lines $L^{(1)}$ and $L^{(2)}$ emanating from the origin at angles $\gamma^{(1)}$ and $\gamma^{(2)}$. These lines define an interior domain with included angle $\omega_1 < \pi$ and an exterior domain with included angle $\omega_2 > \pi$.

designated Ω_1 and Ω_2 respectively. Geometry and the polar coordinate system are illustrated in Fig. 2, where, for convenience, it is assumed that $L^{(1)}$ coincides with the positive x-axis.

The applied electromagnetic field is a TM or TE polarized plane wave incident from exterior wedge Ω_2 onto the vertex at time $t = 0$, with a unit step or jump across the wave front. More conventional time-harmonic response may be obtained by convolving a sinusoid with the solution to this step input. There is no need to specify a particular wave polarization at the outset since only constants in the boundary conditions differ.

Each wedge supports a scalar wave function, f_1 or f_2 , in domains Ω_1 and Ω_2 respectively. Therefore, the two governing wave equations can be written in vector form as

$$(3.1) \quad \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}^2 \Delta \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} - \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}_{tt} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \rightarrow \quad c^2 \Delta f - f_{tt} = 0$$

where $c_1 = 1/\sqrt{\epsilon_1 \mu}$ and $c_2 = 1/\sqrt{\epsilon_2 \mu}$ are wave speeds, scalar Laplacian $\Delta[\] = [\]_{rr} + [\]_{\theta\theta}/r^2$ operates on the elements of the vector and partial derivatives are indicated by r , θ , and t subscripts. Definitions of diagonal matrix c and vector f are clear from (3.1).

The two wave functions are coupled by boundary conditions on set $\partial\Omega$, the union of Ω -domain limit points. In addition to the wedge interface, $L \equiv L^{(1)} \cup L^{(2)}$, these limit points may include infinitesimal circular arcs bounding the vertex as $r \rightarrow 0$. The two boundary conditions on the interface, (2.5), can be written

$$(3.2) \quad \begin{bmatrix} 0 & 0 \\ v_1 & -v_2 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}_{\theta} + \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \rightarrow \quad v f_{\theta} + \tau f_t = 0$$

where definitions of boundary condition matrices v and τ are obvious. Since the matrix coefficients have independent rows, each term in (3.2) can be multiplied by an arbitrary coefficient, e.g., to nondimensionalize or otherwise transform the boundary conditions. Continuity of f_t rather than f itself is included in (3.2) to make the order of differentiation the same in each equation, thereby simplifying the analysis with no loss of generality. For simplicity, no conditions are considered on the vertex, nor are any source terms included on the interface.

Completion of the problem statement requires two initial conditions representing the incident plane wave. These conditions are prescribed on the wave function vector and a derivative in time or space, or two derivatives, provided only that they are independent. Conditions on f_t and f_{θ} are preferred for the advantage of working with the same order of derivative. The incident plane wave in Ω_2 for $t \leq 0$ is given by step function $H(c_2 t - x \cos \phi - y \sin \phi)$, unity for positive argument and zero for negative. Constants $\cos \phi$ and $\sin \phi$ are components of the wave front's unit normal at angle ϕ to the $L^{(1)}$ interface (x-axis). Converting the argument to polar coordinates and dividing by

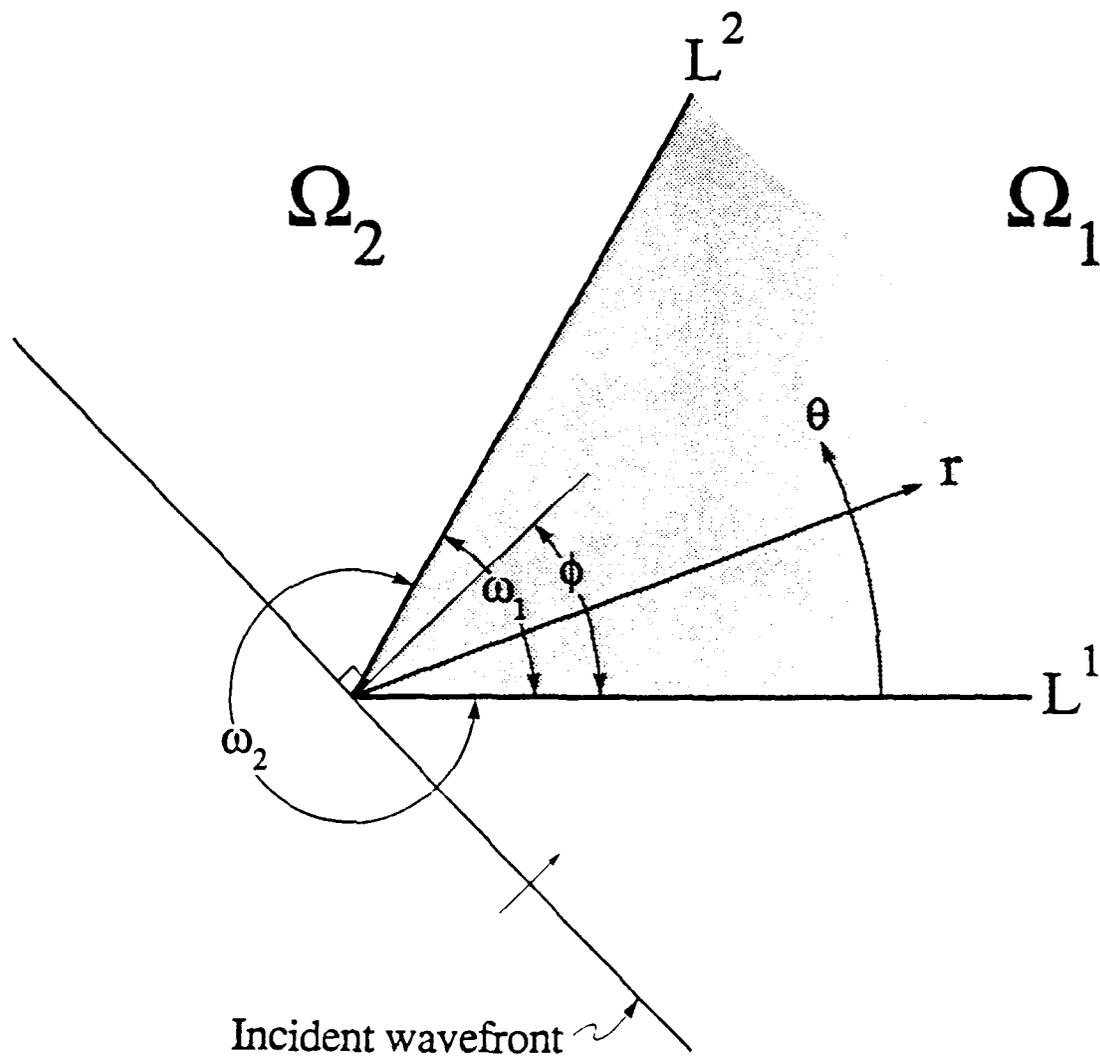


Figure 2. Geometry of the canonical diffraction problem.

r yields

$$(3.3) \quad f_2(r, \theta, t) = H(c_2 t/r - \cos(\theta - \phi))$$

Taking derivatives and setting $t = 0$ gives the initial conditions in terms of the delta function as

$$(3.4) \quad \frac{\partial f_2(r, \theta, 0)}{\partial t} = \frac{c_2}{r} \delta(-\cos(\theta - \phi)) \quad , \quad \frac{\partial f_2(r, \theta, 0)}{\partial \theta} = \sin(\theta - \phi) \delta(-\cos(\theta - \phi))$$

where the argument has two simple zeros in interval $0 \leq \theta \leq 2\pi$, one at $\theta = \phi + \pi/2$ with positive slope, and the other at $\theta = \phi + 3\pi/2$ with negative slope. Based on these slopes, the delta function's argument is further transformed to yield initial condition on wave vector f in the form

$$(3.5) \quad \frac{1}{c_2} f_1(r, \theta, 0) = \pm f_\theta(r, \theta, 0) = \delta(\theta - \phi - \pi \pm \pi/2) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

with the (+) sign for the half of the wave front on $\theta = \phi + \pi/2$, and the (-) sign for the other half on $\theta = \phi + 3\pi/2$.

WAVE SOLUTIONS AND THE INITIAL-BOUNDARY PROBLEM

4. Self-similarity and Homogeneous Vector Solutions

Because the generalized diffractor lacks a characteristic length, self-similar solutions of governing equations (3.2), (3.3), and (3.7) are admissible. Such solutions are given by homogeneous functions of degree h , namely, functions satisfying $f(r, \theta, t) = t^h f(r/t, \theta)$. The most tractable and directly useful for the present case are of degree zero. In principle, functions of non-zero degree can be obtained from these by differentiation or integration.

Defining scalar similarity variable $\rho \equiv r/t$ and transforming vector wave equation (3.1) using differential relation $r f_r = -t f_t = \rho f_\rho$ for homogeneous functions of degree zero, gives the reduced wave equation

$$(4.1) \quad \rho^2 (c^2 - \rho^2 I) f_{\rho\rho} + c^2 f_{\theta\theta} + \rho (c^2 - 2\rho^2 I) f_\rho = 0$$

where I is the identity matrix. Observe that the leading coefficient vanishes at $\rho I = 0$ and $\rho I = c$, indicating a singular point at the vertex and singular circle on the diffracted wave front, respectively. It follows that the reduced wave equation is of mixed type across $\rho I = c$, i.e., hyperbolic outside the diffracted wave front and elliptic inside.

A direct solution of this self-similar form of the vector wave equation begins with the transformation to characteristic coordinates. Writing the characteristic equation, e.g., see [8], and solving for $d\theta/d\rho$ gives the ordinary differential equation

$$(4.2) \quad -\frac{d\theta}{d\rho} = \pm c (\rho \sqrt{\rho^2 I - c^2})^{-1} = \pm \frac{d(\cos^{-1} c/\rho)}{d\rho} \equiv C_\pm$$

which also serves to define the so-called characteristic function, $C_\pm(\rho)$, exhibiting a pole at $\rho = 0$, branch point at $\rho I = c$, and a double zero as $\rho \rightarrow \infty$, with double-valuedness included explicitly by means of the \pm sign. Principal values are assumed for the square root and inverse cosine functions, which are continued through the branch point as

$$(4.3) \quad \sqrt{\rho^2 I - c^2} \rightarrow i \sqrt{c^2 - \rho^2 I} \quad , \quad \cos^{-1} c/\rho \rightarrow i \cosh^{-1} c/\rho$$

These continuations insure that (4.2) is valid for all values of ρ . Multiplying (4.2) by $d\rho$, integrating and rearranging yields

$$(4.4) \quad \theta I \pm \cos^{-1} c/\rho = w_\pm \equiv u \pm v$$

where diagonal matrices $w_\pm = u \pm v$ are constants of integration defining the characteristic coordinates. Since $v \equiv \cos^{-1} c/\rho \rightarrow i \cosh^{-1} c/\rho$, diagonal elements of v are real for $\rho I > c$ and imaginary for $\rho I < c$, hence, the characteristics are real outside the diffracted wave front and complex inside.

Reduced vector wave equation (4.1) is transformed to characteristic coordinates by replacing

derivatives via the chain rule as

$$(4.5) \quad f_{\theta} = f_{w_+} + f_{w_-} \quad , \quad f_{\rho} = C_+ f_{w_+} + C_- f_{w_-}$$

This yields the normal vector form

$$(4.6) \quad f_{w_+ w_-} = 0$$

which can of course be integrated directly giving d'Alembert's classic solution

$$(4.7) \quad f(\rho, \theta) = F_+(w_+) + F_-(w_-)$$

where F_{\pm} are two unknown vector functions. This solution is of mixed type, changing from hyperbolic (real characteristics) to elliptic (complex characteristics) across $v = 0$. Singular line $v = 0$ supports the diffracted wave front and is itself characteristic since it satisfies characteristic equation (4.2), albeit degenerately. In the elliptic domain note that F_{\pm} are complex conjugates and $C_+ = -C_-$ is imaginary. Therefore, in terms of F_{\pm} , derivatives of f become

$$(4.8) \quad f_{\theta} = F_+' + F_-' \quad , \quad f_{\rho} = C_+ F_+' + C_- F_-' = C_+(F_+' - F_-')$$

For later reference it is useful to examine asymptotic solutions of the reduced wave equation near its singularities at $\rho I = 0$ and $\rho I = c$. In the neighborhood of the origin, (4.1) and its solution by separation of variables are

$$(4.9) \quad \rho^2 f_{\rho\rho} + \rho f_{\rho} + f_{\theta\theta} \approx 0 \quad \Rightarrow \quad f(\rho, \theta) = \begin{cases} \ln \rho (a + \theta b) & , \quad \lambda = 0 \\ \rho^{\lambda} (a \cos \theta \lambda + b \sin \theta \lambda) & , \quad \lambda \neq 0 \end{cases}$$

where coefficients a and b , and separation parameter λ are real diagonal matrices. On $L^{(1)}$ and $L^{(2)}$ these must satisfy boundary condition (2.5) with normal coordinate n replaced by θ , where f_1, f_2 are evaluated at $\theta = 0, 2\pi$, respectively, on $L^{(1)}$, and at $\theta = \omega_1$ on $L^{(2)}$. For the $\lambda = 0$ solution this gives $b = 0$ and $a_1 = a_2$. However, for $\lambda \neq 0$ the boundary conditions can only be satisfied if $v_1 = \pm v_2$. The minus sign is irrelevant, while the plus sign corresponds to the TM polarization and yields coefficients $a_1 = a_2, b_1 = b_2$, and exponents $\lambda_1 = \lambda_2 = \pm m$, for m a positive integer. The negative exponent must be eliminated because the resulting energy density, proportional to the square of $f(\rho, \theta)$, is nonintegrable. The positive exponent must also be eliminated because it requires that the field vanish like ρ^m in a zone spreading uniformly from the vertex as time goes to infinity for a constant $\rho \equiv r/t \ll 1$. This null zone is nonphysical and consequently, the only solution valid near the vertex is the first of (4.11), i.e., $f(\rho, \theta) \approx a_1 \ln \rho$.

In the neighborhood of the diffracted wave front, i.e., $\rho I = c \pm \xi$ where $0 \leq \xi \ll I$, reduced wave equation (4.1) and its solution by separation of variables are

$$(4.10) \quad 2\xi f_{\xi\xi} + f_{\xi} \pm \frac{1}{c} f_{\theta\theta} \approx 0 \quad \Rightarrow \quad f(\rho, \theta) \approx (1 + k^{(\pm)} \sqrt{|\rho I - c|}) (a + \theta b)$$

where $k^{(\pm)}, a$, and b are diagonal matrix constants, with $k^{(\pm)}$ defined on either side of the circle; the separation parameter is identically zero. This solution only yields a local estimate of the solution's behavior because boundary conditions are not considered. Nonetheless, (4.10) shows that f and

f_θ are continuous across the circle and have a vertical tangent there by virtue of the $\sqrt{|c - \rho|}$ term. Note that continuity also follows by contradiction since the well-known geometric amplitude decay at a jump or infinity across a cylindrical wave front, namely, $1/\sqrt{r}$, is inconsistent with the self-similarity assumed here. Such jumps would require a homogeneous function of degree $h = -1/2$ rather than $h = 0$. By the same reasoning, continuity applies to all θ derivatives because there is no explicit θ dependency in (4.1).

5. The Vector Initial-Boundary Value Problem

In order to solve the diffraction problem, particular solutions for F_\pm must be found that satisfy the initial and boundary conditions on the transformed wave domain. These conditions must first be transformed to their self-similar, characteristic forms.

Transforming the boundary conditions to the similarity variable gives

$$(5.1) \quad v f_\theta + \tau f_\rho = 0$$

where the second term has been multiplied by $-t^2/r$ to achieve this form. The boundary condition on F_\pm is found by eliminating f_θ and f_ρ in (5.1) using (4.8), giving

$$(5.2) \quad [v + \tau C_+] F_+' + [v + \tau C_-] F_-' = 0$$

or, in component form

$$(5.3) \quad \begin{bmatrix} C_{1+} & -C_{2+} \\ v_1 & -v_2 \end{bmatrix} \begin{bmatrix} F_{1+}' \\ F_{2+}' \end{bmatrix} + \begin{bmatrix} C_{1-} & -C_{2-} \\ v_1 & -v_2 \end{bmatrix} \begin{bmatrix} F_{1-}' \\ F_{2-}' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Transforming derivatives in initial condition (3.5) to the similarity variable, where $tf_t = -\rho f_\rho$, yields

$$(5.4) \quad f_\theta(r, \theta, 0) = \lim_{\rho \rightarrow \infty} f_\theta(\rho, \theta) \quad , \quad f_t(r, \theta, 0) = \lim_{\rho \rightarrow \infty} \frac{-\rho^2}{r} f_\rho(\rho, \theta)$$

Replacing the left sides from (3.5), the right sides from (4.8) with $C_+ \rightarrow c/\rho^2$ as $\rho \rightarrow \infty$, and solving for the limits of F_\pm' gives

$$(5.5) \quad \lim_{\rho \rightarrow \infty} F_\pm' = \mp \delta(u_2 - \phi - \pi \mp \pi/2) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

These transformations of the boundary and initial conditions to self-similar forms (5.2) and (5.5) provide a complete statement of the mixed initial-boundary problem for F_\pm' in the transformed wave domains illustrated in Fig. 3 for an example that will be used later to demonstrate particular solutions. Each of the two mixed domains drawn in the figure, Ω_n , $n = 1, 2$, are spanned by diagonal coordinates u_n and v_n , with hyperbolic Ω_{nH} and elliptic Ω_{nE}

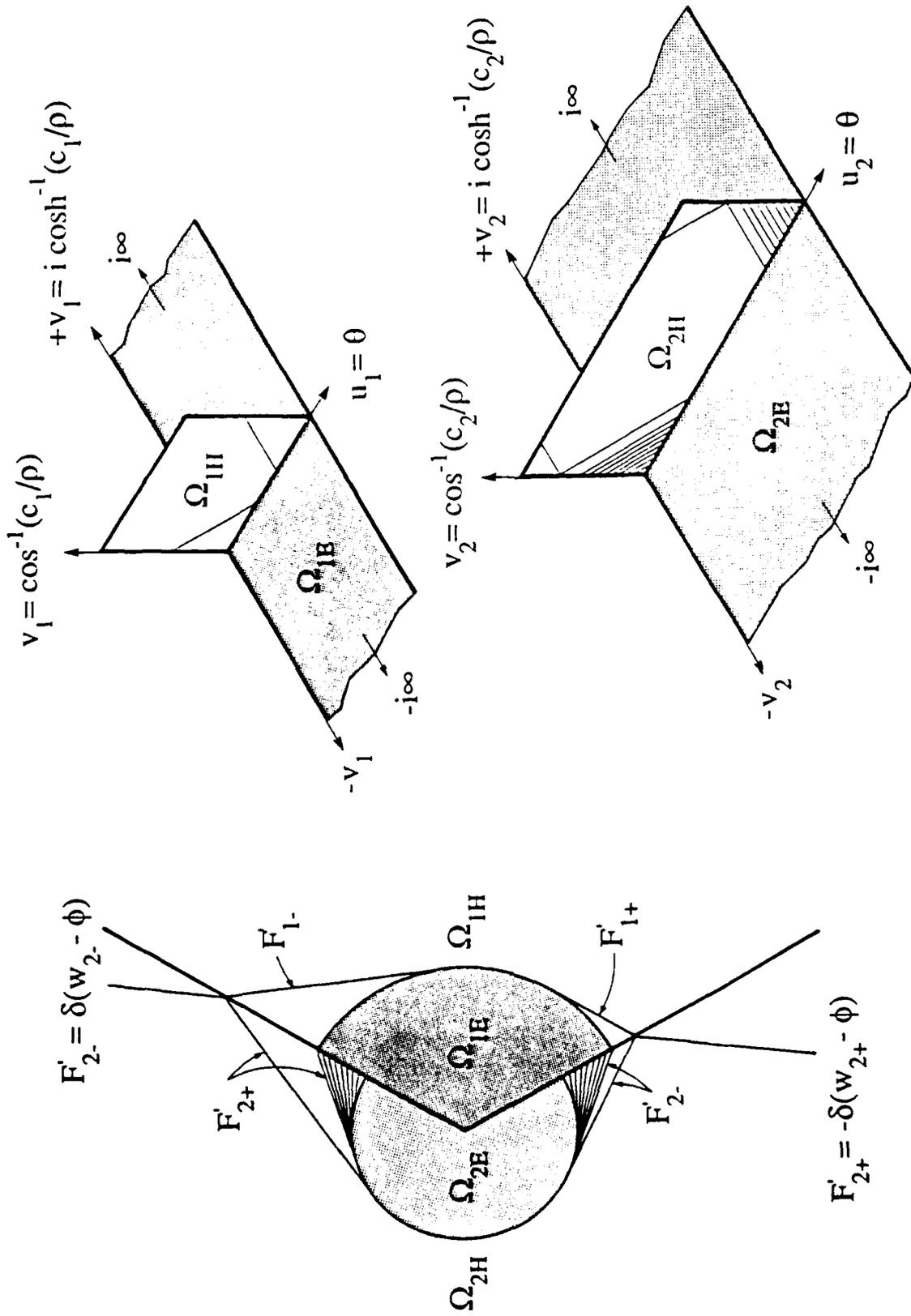


Figure 3. Transformation of the vector wave domain by the self-similar solution.

subdomains corresponding to real and imaginary v_n , respectively. The conjugate elliptic domains covered by complex variables w_{n+} and w_{n-} , are drawn perpendicular to the hyperbolic domains covered by -45° and $+45^\circ$ lines, namely, $w_{n\pm} = u_n \pm v_n$. Limits on u_n are $\gamma^{(n)} \leq u_n \leq \gamma^{(n)} + \omega_n$, where $\gamma^{(n)}$ is the angle of $L^{(n)}$ and ω_n is the included angle of Ω_n . Real limits on v_n are $0 \leq v_n \leq \pi/2$ (hyperbolic subdomain), where $v_n = \pi/2$ corresponds to the initial state at $t = 0$ (or $r \rightarrow \infty$); and $v_n = 0$ is the map of $r = c_n t$, the diffracted wave front. Imaginary limits on v_n are $i0 < v_n < i\infty$, where the upper limit is the map of limiting arcs surrounding the vertex. By virtue of the vector form of the problem, subsequent references to these domains, coordinates, and solutions will drop the subscript or superscript. For consistency, matrices γ and ω are defined as

$$(5.6) \quad \gamma \equiv \begin{bmatrix} \gamma^{(1)} & 0 \\ 0 & \gamma^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \omega_1 \end{bmatrix}, \quad \omega \equiv \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}$$

Initial condition (2.4) is given on line $v = \pi/2$, corresponding to $\rho \rightarrow \infty$ (i.e., $t \rightarrow 0$ or $r \rightarrow \infty$), while boundary condition (2.3) is prescribed on lines $u = \gamma$ and $u = \gamma + \omega$.

6. The Characteristic Mapping

Determining global elliptic solutions of the initial-boundary problem is complicated by geometry of the mixed vector wave domain, consisting of a semi-infinite strip in the u - v plane with real (hyperbolic) and complex (elliptic) characteristic coordinates changing across the diffracted front, $v = 0$, also a characteristic. Applying boundary conditions on the strip perimeter is inconvenient because the boundary parametrization changes at corners and the conditions are of mixed type.

Clearly, the problem would be better posed if it involved only elliptic conditions on a circle or half-plane, i.e., a single-parameter boundary. For a conventional scalar elliptic problem on a strip, it is natural to apply a conformal mapping in order to parametrize the boundary by a single variable. A similar approach is effective in the present case, although the transformation is on a characteristic variable rather than the conventional complex variable. This yields the so-called characteristic mapping, $z(w_{\pm})$.

Consider a succession of mappings of characteristic coordinates w_{\pm} beginning with linear transformation

$$(6.1) \quad \pi \omega^{-1} (w_{\pm} - \gamma)$$

which scales the u coordinate of the vector wave domain (i.e., width of the strip) between 0 and π . Taking the cosine

$$(6.2) \quad \cos (\pi \omega^{-1} (w_{\pm} - \gamma))$$

maps the boundary to the real axis, with w_{+} and w_{-} elliptic subdomains going to the lower and

upper half-planes, respectively, while the entire hyperbolic subdomain maps along characteristics to the real axis between $-I$ and $+I$. Finally, taking the reciprocal,

$$(6.3) \quad z = [\cos(\pi\omega^{-1}(w_{\pm} - \gamma))]^{-1} \equiv x + iy$$

maps the elliptic w_+ subdomain to the upper half-plane, w_- to the lower, and the wedge vertex to the origin. The hyperbolic subdomain now maps to the real axis, on $-I \geq x \geq I$. This is the canonical domain for the vector initial-boundary problem.

Characteristic mapping (6.3) of the mixed domain is a generalization of the familiar Schwarz-Christoffel transformation of semi-infinite strips to half-planes. However, the transformation is much more robust in this context since it maps the entire boundary to the real line and effectively removes the hyperbolic part of the wave domain. The elliptic part of the strip maps conformally to the half-plane, while the hyperbolic part collapses or folds nonconformally along the characteristics to a finite or semi-infinite segment of the real axis. This segment is coincident with the map of the $v = 0$ diffracted front. Consequently, boundary conditions on the mixed interfaces are transformed along characteristics to conditions on the diffracted front. There is a one-to-one mapping of points on the diffracted front to $-I \geq x \geq I$, but a multivalued mapping of the hyperbolic and mixed interfaces onto this range.

What remains is a boundary problem for the elliptic part of F_{\pm} on the real axis. This elliptic part is redefined for convenience as

$$(6.4) \quad F_{\pm}(w_{\pm}(z)) \equiv W(z) = U(x,y) + iV(x,y)$$

where $W(z)$ satisfies the reflection principle, i.e., $W(z^*) = W^*(z)$, since F_+ and F_- are complex conjugates and map to the upper and lower half-planes respectively. The problem is to determine $W(z)$ by continuation of the transformed boundary conditions on $y = 0$.

REDUCTION TO ELLIPTIC BOUNDARY CONDITIONS

7. Hyperbolic Solutions

It is necessary to begin by solving the initial part of the problem, i.e., solutions in the hyperbolic subdomains. These subdomains consist of rectangles in the w plane and the solution procedure is simply the method of characteristics for the 1-D wave equation, e.g., the string equation. However, in this problem there are two types of hyperbolic unknowns to be determined—non-mixed and mixed. Non-mixed or explicit unknowns are associated with the initial condition and constitute the system of incident, reflected, and transmitted plane waves. Mixed or implicit unknowns are associated with the Mach envelopes in the slower wedge and are composed of plane waves excited by the “superluminal” diffracted wave in the faster wedge grazing the interface. The hyperbolic wave solutions are developed in this section, and demonstrated for the example in Fig. 3.

The incident plane wave follows from initial condition (5.5) on the strip end by direct continuation. Since $v \rightarrow \pi/2I$ as $t \rightarrow 0$ or $\rho \rightarrow \infty$, and therefore $w_{2\pm} \rightarrow u_2 \pm \pi/2$, then (5.5) may be rewritten as

$$(7.1) \quad F_{2+}' = -\delta(w_{2+} - (\phi + 2\pi)) \quad , \quad F_{2-}' = \delta(w_{2-} - \phi)$$

Characteristics $w_{2+} = \phi + 2\pi$, $w_{2-} = \phi$ support the two halves of the incident wave front on either side of the interior wedge—with the plus characteristic below and the minus characteristic above. Moving along the characteristics toward the interior wedge, the waves either reflect off each interface or one may terminate directly on the exterior diffracted front in the case of a shadow. The case depends on whether the wave front normal through the vertex at $t = 0$ is inside or outside the interior wedge, e.g., see Fig. 2. Reflection points at $\rho^{(1)}$ on $L^{(1)}$ and $\rho^{(2)}$ on $L^{(2)}$, if they exist, follow by setting the argument of the delta function in (7.1) to zero and solving for ρ , whence

$$(7.2) \quad \rho^{(1)} = c_2/\cos(\phi) \quad , \quad \rho^{(2)} = c_2/\cos(\omega_1 - \phi)$$

Boundary condition (5.2) must be solved for the reflected and transmitted wave coefficients at these points.

If the reflection point is on a hyperbolic interface, i.e., $\rho^{(i)} > \max.[c_1, c_2]$, then the boundary condition can be solved explicitly, with solutions continued off the boundaries as reflected waves in the exterior and transmitted waves in the interior. Continuation of the delta function is accomplished using the identity

$$(7.3) \quad g'(s_0) \delta(g(s)) \equiv \delta(s - s_0)$$

where s_0 is a simple zero of $g(s)$. The complete system of non-mixed waves generated by reflection and transmission of the incident wave becomes quite involved if the interior wedge angle is small. However, a straightforward tree-type data structure of characteristics organizes the

system for complete analysis. Thus, the non-mixed hyperbolic problem reduces to systematically calculating the reflection and transmission coefficients from (5.2) at each non-mixed branch point on the characteristic tree until these characteristics terminate on a mixed interface or the diffracted front.

If the reflection point of the incident wave, or subsequent reflected and transmitted waves, is on a mixed interface, i.e., $\max.[c_1, c_2] > \rho^{(i)} > \min.[c_1, c_2]$, then reflection occurs within a Mach envelope consisting of a continuum of mixed waves emanating from the interface. No plane wave is transmitted since the solution there is elliptic (so-called evanescent). Furthermore, the reflected wave can only be solved implicitly in terms of this elliptic solution because boundary condition (5.2) is mixed (hyperbolic-elliptic). As the interior wedge angle decreases, these Mach envelopes intersect and eventually reflect between interfaces. This does not affect solvability but does make the analysis much more tedious. Again, a tree-type data structure is helpful in organizing the system.

To illustrate these points, explicit and implicit hyperbolic solutions will be determined for the case shown in Fig. 3. Referring to the figure, the $w_{2-} = \phi$ characteristic supporting incident wave F_{2-}' above the interior wedge contacts $L^{(2)}$ at $\rho^{(2)}$ on a hyperbolic segment. The unknown reflected and transmitted waves are F_{2+}' and F_{1-}' respectively, and F_{1+}' vanishes since there is no incident interior wave. Similarly, the $w_{2+} = \phi + 2\pi$ characteristic supporting incident wave F_{2+}' below the interior wedge contacts $L^{(1)}$ at $\rho^{(1)}$, also on a hyperbolic segment, where F_{2-}' and F_{1+}' are reflected and transmitted waves and F_{1-}' vanishes. Accordingly, solving boundary condition (5.3) on either interface yields

$$(7.4) \quad F_{2\mp}' = -A_R^{(i)} F_{2\pm}'^{(i)} \quad , \quad F_{1\pm}' = \frac{v_2/v_1}{Q^{(i)}} A_T^{(i)} F_{2\pm}'^{(i)}$$

where the upper and lower signs correspond to solutions on $L^{(i)}$, $i = 1, 2$ respectively, $F_{2\pm}'^{(i)}$ is the incident wave, (7.1) evaluated on $L^{(i)}$. The other terms are given by

$$(7.5) \quad A_R^{(i)} = A_T^{(i)} - 1 \equiv \frac{Q^{(i)} - 1}{Q^{(i)} + 1} \quad , \quad Q(\rho) \equiv \frac{v_2 |C_{1+}(\rho)|}{v_1 |C_{2+}(\rho)|}$$

with $Q^{(i)} \equiv Q(\rho^{(i)})$, and $A_R^{(i)}$, $A_T^{(i)}$ defining the conventional Fresnel reflection and transmission coefficients.

These boundary solutions are continued off the interfaces into Ω_1 and Ω_2 by changing the delta function's argument in (7.1) at $\rho^{(i)}$ according to (7.3). Thus, the reflected and transmitted waves become

$$(7.6) \quad F_{2\mp}'(w_{2\mp}) = \pm A_R^{(i)} \delta(w_{2\mp} - w_2^{(i)}) \quad , \quad F_{1\pm}'(w_{1\pm}) = \mp A_T^{(i)} \delta(w_{1\pm} - w_1^{(i)})$$

$$w_2^{(1)} = 2\pi - \phi, \quad w_2^{(2)} = 2\omega_1 - \phi, \quad w_1^{(1)} = \cos^{-1}(c_1/\rho^{(1)}), \quad w_1^{(2)} = \omega_1 - \cos^{-1}(c_1/\rho^{(2)})$$

As an example of solutions on a mixed interface, consider Fig. 3 when $c_1 > \rho^{(i)} > c_2$. In boundary condition (5.3), $F_{1\pm}'$ are now elliptic while $C_{1\pm}$ are imaginary. Solving on either interface for the reflected wave yields

$$(7.8) \quad F_{2\mp}' = \pm \frac{v_1}{v_2} Q i (F_{1+}' - F_{1-}') + F_{2\pm}^{(i)} = \frac{v_1}{v_2} (F_{1+}' + F_{1-}') - F_{2\pm}^{(i)}$$

where again, $F_{2\pm}^{(i)}$ on the right is given by (7.1) on $L^{(i)}$ and the upper and lower signs correspond to solutions on $L^{(1)}$ and $L^{(2)}$. These hyperbolic boundary solutions are real since $F_{1\pm}'$ are complex conjugates. Thus, the mixed waves in the Mach envelope can only be solved implicitly in terms of the elliptic solutions plus the incident wave. Observe that the reflected wave can be eliminated from (7.8) resulting in

$$(7.9) \quad (1 \mp iQ) F_{1+}' + (1 \pm iQ) F_{1-}' = 2 \frac{v_2}{v_1} F_{2\pm}^{(i)}$$

This provides a scalar condition on the elliptic unknown along the mixed interface that can, in principle, be used to determine the local behavior of $F_{1\pm}'$ and thereby evaluate (7.8) at the reflection point. However, this will be deferred to Sec. 10 where global elliptic solutions are considered.

8. Elliptic Boundary Conditions

Since the characteristic transformation is only conformal on the non-mixed part of the boundary ($F_{1\pm}'$ and $F_{2\pm}'$ both elliptic) the form of boundary condition (5.2) is invariant there and nowhere else. In contrast, whenever F_{\pm}' is mixed ($F_{1\pm}'$ elliptic and $F_{2\pm}'$ hyperbolic, or vice versa) the boundary condition must be modified by eliminating the hyperbolic solutions, resulting in elliptic conditions only. Thus the simplified geometry provided by the characteristic mapping is obtained at the expense of a modified boundary condition on $-I \geq x \geq I$.

Conditions on the non-mixed elliptic part of the boundary, which extends from the origin to the map of $\rho = \min.[c_1, c_2]$ are simply found by substituting (6.4) into (5.2) and rewriting as

$$(8.1) \quad C_{1+}(W_1^+ - W_1^-) - C_{2+}(W_2^+ - W_2^-) = 0$$

$$v_1(W_1^+ + W_1^-) - v_2(W_2^+ + W_2^-) = 0$$

where \pm superscripts denote limits from above and below the real axis, hence complex conjugates. Solving for the real and imaginary parts gives

$$(8.2) \quad v_2 = \frac{v_1}{v_2} Q v_1, \quad v_1 U_1 = v_2 U_2$$

where the relation between ρ in $Q(\rho)$, x_1 , and x_2 is readily found from (6.3) to be

$$(8.3) \quad \frac{1}{\rho} = \frac{1}{c_1} \cosh\left[\frac{\omega_1}{\pi} \cosh^{-1}\left|\frac{1}{x_1}\right|\right] = \frac{1}{c_2} \cosh\left[\frac{\omega_2}{\pi} \cosh^{-1}\left|\frac{1}{x_2}\right|\right]$$

Note that these elliptic boundary conditions yield another, more explicit interpretation. Effectively, (5.2) states that $[v + \tau C_{\pm}]W^{\pm}$ is imaginary on the boundary, equal to $\pm iR$ say, whence, solving for W^{\pm} and expanding gives the factorization

$$(8.4) \quad W^{\pm} = \frac{\sqrt{c_1^2 - \rho^2} \sqrt{c_2^2 - \rho^2}}{v_1 c_2 \sqrt{c_1^2 - \rho^2} - v_2 c_1 \sqrt{c_2^2 - \rho^2}} \begin{bmatrix} v_2 \rho R_1 \pm \frac{i c_2}{\sqrt{c_2^2 - \rho^2}} R_2 \\ v_1 \rho R_1 \pm \frac{i c_1}{\sqrt{c_1^2 - \rho^2}} P_2 \end{bmatrix}$$

i.e., an explicit expression for boundary values of $W(z)$ in terms of unknown real function $R(x)$. This indicates the singular endpoint behavior expected in W^{\pm} , and will be useful later in establishing the final form of $W(z)$.

Conditions on the mixed hyperbolic-elliptic parts of the boundary follow by eliminating the hyperbolic part. Critical to this elimination is wave field continuity across the diffracted wave front, also a characteristic, since the continuity condition provides an additional equation that can be solved for the hyperbolic solution. Therefore, the hyperbolic solution is eliminated between the boundary and continuity conditions. Continuity of f and f_{θ} has already been established from wave front solution (4.13) and the discussion following it. Therefore, from (4.9), $f_{\theta} = F_+' + F_-'$ is continuous, hence

$$(8.4) \quad (F_+' + F_-')|_{v=0_+} = (F_+' + F_-')|_{v=i0}$$

where the left side is hyperbolic and the right side is elliptic and equal to $2\text{Re}[F_+']$. Effectively, the hyperbolic solutions are eliminated by continuing them to the diffracted wave front and substituting into (8.4), yielding a condition on the real part of the elliptic unknown. For the explicit waves this provides a relation on the elliptic unknown in one domain, however, for the implicit waves, i.e., Mach envelopes, it yields a relation between elliptic unknowns in both domains. In either case the resulting elliptic equations are mapped conformally to the half-plane boundaries by the characteristic transformation, yielding the necessary boundary conditions.

To illustrate the procedure, the example in Fig. 3 will again be used. The characteristic transformation for this case is shown in Fig. 4. Significant points on the interface and diffracted fronts are labeled A through F, where points A' and C' correspond to the tangent point of the leading Mach wave emanating from A and C. Points B and E bisect the diffracted fronts and are mapped to infinity by the transformation. Note that this case, with $c_1 > c_2$, provides the simplest domain in which to formulate and solve the elliptic boundary problem since the hyperbolic solutions (plane waves) do not undergo multiple reflections.

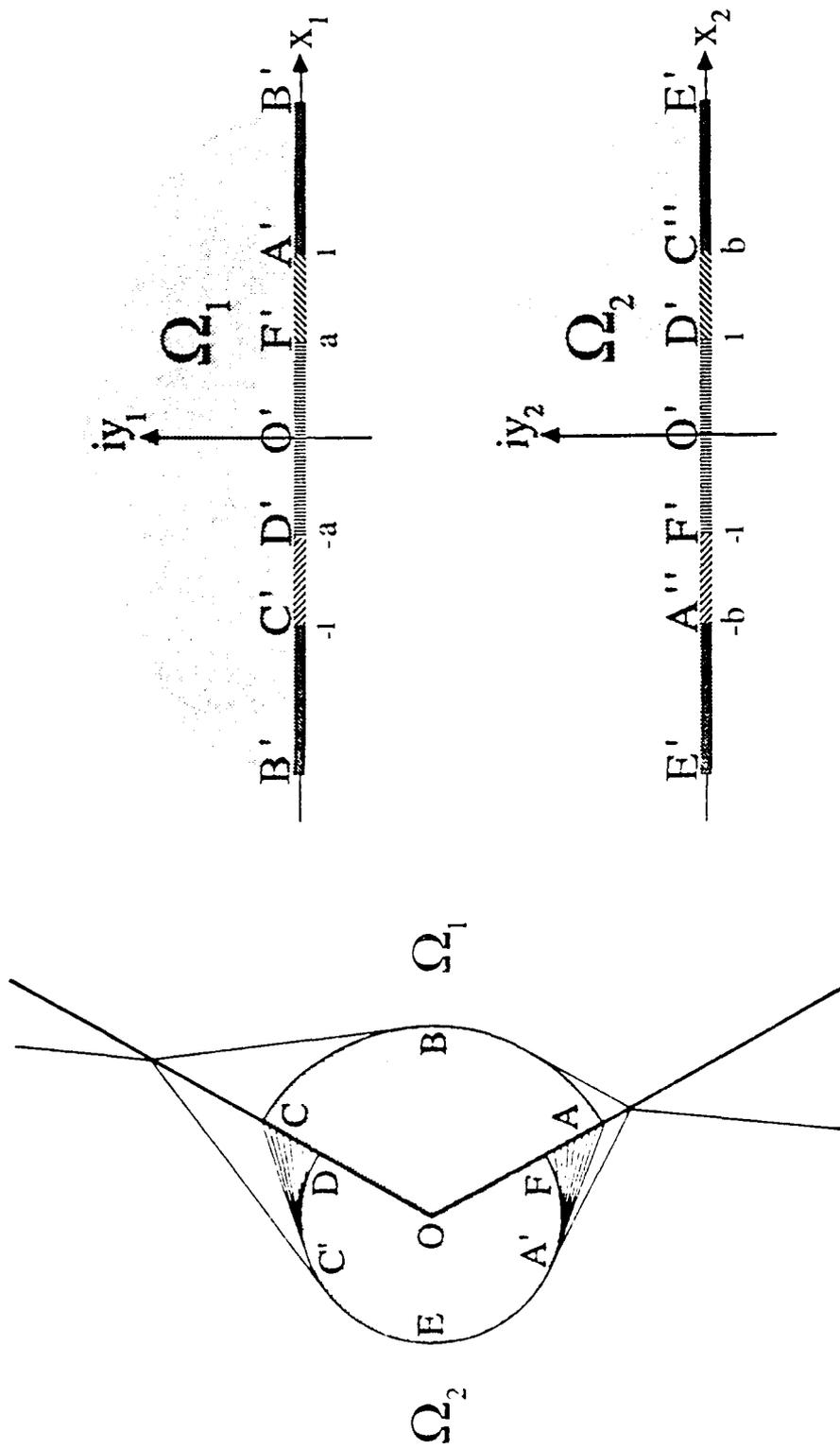


Figure 4. Transformation of the elliptic domains by the characteristic mapping.

For the explicit hyperbolic solutions, continuing the reflected and transmitted solutions in (7.7) along characteristics to their respective diffracted fronts and substituting into (8.4) yields conditions on the elliptic solutions defined there. Characteristic transformation (6.3) then maps the resulting elliptic conditions conformally to the real axis of the z_1 and z_2 half-planes. Transforming the delta functions in (7.7) on the diffracted front via (7.3), substituting into the left side of (8.4), and replacing the right side by $2U = W^+ + W^-$ gives the desired relations as

$$(8.5) \quad U_1 = \frac{1}{2} A_T^{(i)} \zeta_1^{(i)} \delta(x_1 - x_1^{(i)}) \quad , \quad U_2 = \frac{1}{2} A_R^{(i)} \zeta_2^{(i)} \delta(x_2 - x_2^{(i)})$$

$$\zeta^{(i)} \equiv -\pi \omega^{-1} x^{(i)} \sqrt{(x^{(i)})^2 - 1} \quad , \quad x^{(i)} = [\cos \pi \omega^{-1} (\omega^{(i)} - \gamma)]^{-1}$$

valid on the real axes of the complex half-planes, with $i = 1, 2$ on maps of AB and BC in the z_1 plane and A'E and EC' in the z_2 plane, respectively.

To illustrate elimination of the implicit hyperbolic solutions, consider Mach fan solution (7.8). Observe that in this example the Mach fans are supported by only one characteristic and yield so-called single wave zones near the diffracted front. Therefore, continuation along either w_+ or w_- characteristics to the front and substitution into (8.4) yields the elliptic conditions. Hence, on maps of mixed interface segments CD and AF, the elliptic conditions equivalent to (7.8) are

$$(8.6) \quad U_2 = \mp \frac{V_1}{V_2} Q V_1 + \frac{1}{2} F_{2\pm}^{(i)} = \frac{V_1}{V_2} U_1 - \frac{1}{2} F_{2\pm}^{(i)}$$

relating the real part of W_2 to the real or imaginary part of W_1 , with $F_{2\pm}^{(i)}$ representing delta functions (7.1) evaluated on $L^{(i)}$. Transforming the delta function's argument from ρ to x_1 via (7.3) and eliminating U_2 from (8.6) gives

$$(8.7) \quad U_1 \pm Q V_1 = Q^{(i)} \zeta_1^{(i)} \delta(x_1 - x_1^{(i)})$$

$$\zeta_1^{(i)} \equiv -\frac{\pi}{\omega_1} x_1^{(i)} \sqrt{1 - (x_1^{(i)})^2} \quad , \quad x_1^{(i)} = [\cosh \frac{\pi}{\omega_1} \cosh^{-1} (\frac{c_1}{c_2} \cos(\phi - \gamma^{(i)}))]^{-1}$$

which is equivalent to (7.9) and relates the real and imaginary parts of W_1 . The upper and lower signs are for maps of segments AF on $L^{(1)}$ and DC on $L^{(2)}$, respectively, in Fig. 4. In general, the relation between ρ , x_1 , and x_2 follows from (6.3) as

$$(8.12) \quad \frac{1}{\rho} = \frac{1}{c_1} \cosh \left[\frac{\omega_1}{\pi} \cosh^{-1} \left| \frac{1}{x_1} \right| \right] = \frac{1}{c_2} \cos \left[\frac{\omega_2}{\pi} \cos^{-1} \left| \frac{1}{x_2} \right| \right]$$

Thus, for a given x_1 or x_2 , the value of ρ in $Q(\rho)$ and other terms evaluated on the mixed interface are known.

INTEGRAL REPRESENTATIONS AND SINGULAR EQUATIONS

9. The Scalar Riemann-Hilbert Problems

The boundary conditions derived above, namely, (8.2), (8.5), (8.8), and (8.9), relate real and/or imaginary parts of the unknown elliptic functions. These equations couple parts either between the two domains or within one domain only, defining respectively, vector or scalar boundary problems. The scalar problems given by (8.5) and (8.9) are solved below in closed-form. Their solutions establish the fundamental structure of representations used to formulate vector boundary problems (8.2) and (8.8) in the next section.

Determination of an analytic function in the complex plane given a linear relation between its real and imaginary parts on a curve or contour has come to be called the Riemann-Hilbert problem, e.g., see [1], or [2] for an exhaustive treatment. This is precisely the type of boundary problem derived above on maps of mixed segments CD and FA, as well as diffracted fronts ABC and A'EC' on the real axes in Fig. 4. The solution of Riemann-Hilbert problems on the real axis follows a straightforward procedure based on factorization methods and Cauchy-type line integrals. Fundamental to this procedure is the representation of analytic functions in the half-plane by Cauchy-type line integrals over segments of the real axis.

The basis for all that follows is: given the imaginary part of an analytic function on a segment of the real axis, then the function can be represented by a Cauchy-type line integral on this imaginary part over the segment. For example, if $M(z)$ is analytic off the real axis, satisfies the reflection principle, and its imaginary part is known on $[x_a, x_b]$, then representations of the function and its boundary values are

$$M(z) = \frac{1}{\pi} \int_{x_a}^{x_b} \frac{\text{Im}[M^+(s)]}{s-z} ds + N(z) \quad (9.1)$$

$$M^\pm(x) = \frac{1}{\pi} \int_{x_a}^{x_b} \frac{\text{Im}[M^+(s)]}{s-x} ds + N(x) \pm i \text{Im}[M^+(x)]$$

The integral is a particular solution of inhomogeneous equation $M^+ - M^- = i2\text{Im}[M^+]$, while $N(z)$ is a solution of homogeneous equation $N^+ - N^- = 0$, hence, N is real on $[x_a, x_b]$. This type of representation is a generalization of the Poisson integral formula on the half-plane, or equivalently, the Hilbert transform, and is readily mapped to arbitrary curves and contours. The second equation in (9.1) is the limit of the first as $z \rightarrow x$ from above (+) or below (-) the axis. The singular integral is evaluated as a Cauchy principal value and the imaginary part is the result of integrating around the Cauchy pole on an infinitesimal half-circle.

With this background it is natural to begin the analysis with (8.9) after putting it into the so-called Riemann form, i.e., in terms of W_1^\pm as

$$(9.2) \quad (1 \mp iQ) W_1^+ + (1 \pm iQ) W_1^- = 2Q \zeta_1^{(i)} \delta(x_1 - x_1^{(i)})$$

Dividing through by the first coefficient on the left, introducing the factorizations

$$(9.3) \quad \frac{G^+}{G^-} \equiv \frac{1 \pm iQ}{1 \mp iQ} \quad , \quad \frac{B^+}{B^-} \equiv -1$$

and dividing by G^+ and $B^+ = -B^-$ yields

$$(9.4) \quad \frac{W_1^+}{B_1^+ G_1^+} - \frac{W_1^-}{B_1^- G_1^-} = \frac{2Q^{(i)}}{B_1^+ G_1^+ (1 \mp iQ^{(i)})} \zeta_1^{(i)} \delta(x_1 - x_1^{(i)})$$

This is the normal form of Riemann-Hilbert problem (8.9). Comparing to (9.1) the solution is

$$(9.5) \quad W_1(z_1) = B_1(z_1) G_1(z_1) \{H_1(z_1) + P_1(z_1)\}$$

where $H_1(z_1)$ is some homogeneous solution, real on segments $(-1, -a)$, $(a, 1)$, i.e., excluding endpoints, and $P_1(z_1)$ is the particular solution

$$(9.6) \quad P_1 = \left\{ \frac{i}{\pi B_1^+ G_1^+} \frac{Q}{1-iQ} \zeta_1 \right\}^{(1)} \frac{1}{z_1 - x_1^{(1)}} + \left\{ \frac{i}{\pi B_1^+ G_1^+} \frac{Q}{1+iQ} \zeta_1 \right\}^{(2)} \frac{1}{z_1 - x_1^{(2)}}$$

consisting of simple poles. These poles result from integrating the delta function and evaluating via the sifting property. The coefficients are real since B_1^+ is imaginary and G_1^+ is the complex conjugate of $(1 - iQ)$ on $[a, 1]$ and $(1 + iQ)$ on $[-1, -a]$.

The completion of representation (9.5) requires expressions for $B_1(z_1)$ and $G_1(z_1)$. Taking the principal-value logarithm of the factorizations in (9.3) yields the log Riemann-Hilbert forms

$$(9.7) \quad \ln G_1^+ - \ln G_1^- = \ln \frac{1 \pm iQ}{1 \mp iQ} = \pm i 2 \tan^{-1} Q \quad , \quad \ln B_1^+ - \ln B_1^- = \ln(-1) = i\pi$$

A particular solution for $\ln G_1(z_1)$ is therefore

$$(9.8) \quad \begin{aligned} \ln G_1(z_1) &= \frac{1}{\pi} \int_a^1 \frac{\tan^{-1} Q}{s_1 - z_1} ds_1 - \frac{1}{\pi} \int_{-1}^{-a} \frac{\tan^{-1} Q}{s_1 - z_1} ds_1 \\ &= \frac{2}{\pi} \int_a^1 \frac{s_1 \tan^{-1} Q(s_1)}{s_1^2 - z_1^2} ds_1 \equiv \Gamma(z_1) \end{aligned}$$

and a general solution for $\ln B_1(z_1)$ is

$$(9.9) \quad \ln B_1(z_1) = \frac{1}{\pi} \int_a^1 \frac{\pi/2}{s_1 - z_1} ds_1 + \frac{1}{\pi} \int_{-1}^{-a} \frac{\pi/2}{s_1 - z_1} ds_1 + \ln \beta_1(z_1)$$

$$= \ln \sqrt{\frac{(z_1 + a)(z_1 - 1)}{(z_1 - a)(z_1 + 1)}} + \ln \beta_1(z_1)$$

where $\ln \beta_1$ is the homogeneous solution. Exponentials of (9.8) and (9.9) yield

$$(9.10) \quad G_1(z_1) = e^{\Gamma(z_1)} \quad , \quad B_1(z_1) = \beta_1(z_1) \sqrt{\frac{(z_1 + a)(z_1 - 1)}{(z_1 - a)(z_1 + 1)}}$$

Thus, B_1 has branch cuts on $[-1, -a]$ and $[a, 1]$, and since $\beta_1(z_1)$ may have arbitrary zeros or poles at the endpoints, $B_1(z_1)$ can exhibit square root branch points in either the numerator or the denominator, despite the ordering in (9.10).

The remaining scalar Riemann-Hilbert problem in the z_1 plane is given by the first of (8.5), representing the transmitted waves on A'B'C' in Fig. 4. The Riemann form is

$$(9.11) \quad W_1^+ + W_1^- = A_T^{(1)} \zeta_1^{(1)} \delta(x_1 - x_1^{(1)})$$

These delta functions are continuations of those in (8.9) as the reflection points move from mixed to non-mixed segments of the interface. Therefore, the solution of (8.5) may be included directly in that of (8.9), namely, (9.2), by extending the definitions of B_1 , H_1 , and P_1 . In particular, if the branch points in B_1 at $x = \pm 1$ are moved off to $\pm\infty$ so that $B_1(z)$ has branch cuts on $[a, \infty)$, $(-\infty, -a]$ rather than $[a, 1]$, $[-1, -a]$, then dividing (9.10) by $G_1^+ B_1^+ = -G_1^- B_1^-$ and solving the Riemann-Hilbert problem yields (9.5) with P_1 redefined as

$$(9.12) \quad P_1(z_1) = \left\{ \frac{i A_T \zeta_1}{2\pi B_1^+ G_1^+} \right\}^{(1)} \frac{1}{z_1 - x_1^{(1)}} + \left\{ \frac{i A_T \zeta_1}{2\pi B_1^+ G_1^+} \right\}^{(2)} \frac{1}{z_1 - x_1^{(2)}}$$

which is real since B_1^+ is imaginary and G_1^+ is real on $[1, \infty), (-\infty, -1]$.

In the z_2 plane, the scalar Riemann-Hilbert problem is given by the second of (8.5) representing reflected waves on FE'D' in Fig. 4. Following the above procedure, the normal form can be written

$$(9.13) \quad \frac{W_2^+}{B_2^+} - \frac{W_2^-}{B_2^-} = \frac{A_R^{(1)} \zeta_2^{(1)}}{B_2^+} \delta(z_2 - x_2^{(1)})$$

where $B_2(z_2)$ has branch cuts on $(-\infty, -b]$, $[b, \infty)$, i.e.,

$$(9.14) \quad B_2(z_2) = \beta_2(z_2) \sqrt{\frac{z_2 + b}{z_2 - b}}$$

where β_2 has poles or zeros at the endpoints. A general solution is

$$(9.15) \quad W_2(z_2) = B_2(z_2) \{H_2(z_2) + P_2(z_2)\}$$

where homogeneous solution H_2 is real on C'E'A" and particular solution P_2 is

$$(9.16) \quad P_2(z_2) = \left\{ \frac{iA_R \zeta_2}{2\pi B_2^+} \right\}^{(1)} \frac{1}{z_2 - x_2^{(1)}} + \left\{ \frac{iA_R \zeta_2}{2\pi B_2^+} \right\}^{(2)} \frac{1}{z_2 - x_2^{(2)}}$$

which is real except at the poles. Note that this solution does not include the case where the point of reflection is on the mixed interface segment since the corresponding boundary condition, (8.8), is a vector Riemann-Hilbert problem.

10. The Vector Riemann-Hilbert Problems

Solutions of the scalar Riemann-Hilbert problems, (9.5) and (9.15), yield global representations of analytic functions $W_1(z_1)$ and $W_2(z_2)$ that include unknown functions $H_1(z_1)$ and $H_2(z_2)$. To complete the solution, these unknowns must be determined so as to satisfy the remaining vector Riemann-Hilbert problems, given by (8.2) on maps of DOF and (8.8) on maps of C'D and A'F, namely

$$(10.1) \quad V_1 = \frac{v_2}{v_1} \frac{1}{Q} V_2, \quad U_1 = \frac{v_2}{v_1} U_2$$

for x_1 on $[-a, a]$ and x_2 on $[-1, 1]$, and

$$(10.2) \quad U_1 = \frac{v_2}{v_1} U_2 + \frac{1}{2} Q^{(i)} \zeta_1^{(i)} \delta(x_1 - x_1^{(i)})$$

for x_1 on $[-1, -a]$, $[a, 1]$ and x_2 on $[-b, -1]$, $[1, b]$.

Homogeneous factors H_1 and H_2 are, effectively, residuals derived by multiplicative and additive removal of the scalar boundary behavior. It follows that the imaginary part of H_1^+ exists only on $[-a, a]$, i.e., H_1^+ is complex on $[-a, a]$ and real elsewhere. Similarly, the imaginary part of H_{2+} exists only on $[-b, b]$. Therefore, from (9.1), representations for H_1 and H_2 can be written in terms of Cauchy integrals on the imaginary part, plus homogeneous solutions. Accordingly, replacing H_1 in (9.5) and H_2 in (9.15) yields the final representations for $W_1(z_1)$ and $W_2(z_2)$ as

$$(10.3) \quad W_1(z_1) = B_1(z_1) G_1(z_1) \left\{ \frac{1}{\pi} \int_{-a}^a \frac{(B_1 G_1)^{-1} K_1(s_1)}{s_1 - z_1} ds_1 + E_1(z_1) + P_1(z_1) \right\}$$

$$(10.4) \quad W_2(z_2) = B_2(z_2) \left\{ \frac{1}{\pi} \int_{-b}^b \frac{(B_2)^{-1} K_2(s_2)}{s_2 - z_2} ds_2 + E_2(z_2) + P_2(z_2) \right\}$$

where $\text{Im}[H_1] \equiv (B_1 G_1)^{-1} K_1$ and $\text{Im}[H_2] \equiv (B_2)^{-1} K_2$ in the Cauchy integrals, K_1 and K_2 are unknown density functions, and homogeneous solutions E_1 and E_2 are real on the real axes, hence, entire functions. Taking the limit as $z_1 \rightarrow x_1$ gives boundary values of W_1 on $[-a, a]$ in the form

$$(10.5) \quad W_1^+ = B_1 G_1 \left\{ \frac{1}{\pi} \int_{-a}^a \frac{(B_1 G_1)^{-1} K_1(s_1)}{s_1 - x_1} ds_1 + E_1 \right\} + B_1 G_1 P_1 + i K_1$$

where the integral is a principal value. The imaginary part of this expression is K_1 , hence, $K_1 = V_1$ on segment $[-a, a]$. On $[-1, -a]$, $[a, 1]$ the boundary values are

$$(10.6) \quad W_1^+ = B_1^+ G_1^+ \left\{ \frac{1}{\pi} \int_{-a}^a \frac{(B_1 G_1)^{-1} K_1(s_1)}{s_1 - x_1} ds_1 + E_1 \right\} + B_1^+ G_1^+ P_1^+$$

where the integral is real and nonsingular since x_1 is outside the integration interval. Similarly, taking the limit as $z_2 \rightarrow x_2$ gives boundary values of W_2 on $[-b, b]$ as

$$(10.7) \quad W_2^+ = B_2 \left\{ \frac{1}{\pi} \int_{-b}^b \frac{(B_2)^{-1} K_2(s_2)}{s_2 - x_2} ds_2 + E_2 \right\} + B_2 P_2^+ + i K_2$$

On segment $[-1, 1]$ the imaginary part of these boundary values is K_2 , hence, $K_2 = V_2$ there. Note that \pm superscripts are used here when limiting values of a function from above or below the axis have, or may have, different values—on a branch cut or pole in particular.

Given these representations for boundary values of W_1 and W_2 , the first part of the vector Riemann-Hilbert problem in (10.1) for x_1 on $[-a, a]$ and x_2 on $[-1, 1]$ is satisfied by setting $K_1 = V_1$, $K_2 = V_2$ and replacing K_1 in integral representations for W_1 . The second part of this vector problem is satisfied by substituting real parts from (10.5) and (10.7), giving the integral equation

$$(10.8) \quad B_1 G_1 \int_{-a}^a \frac{(B_1 G_1 Q)^{-1} K_2(x_2(s_1))}{s_1 - x_1} ds_1 - B_2 \int_{-1}^1 \frac{(B_2)^{-1} K_2(s_2)}{s_2 - x_2} ds_2 = S$$

$$(10.9) \quad S = \pi(B_2 \{P_2 + E_2\} - \frac{V_1}{V_2} B_1 G_1 \{P_1 + E_1\})$$

The remaining vector problem, (10.2), for x_1 on $[-1, -a]$, $[a, 1]$ and x_2 on $[-b, -1]$, $[1, b]$ yields

$$(10.10) \quad \text{Re}[B_1^+ G_1^+] \int_{-a}^a \frac{(B_1 G_1 Q)^{-1} K_2(x_2(s_1))}{s_1 - x_1} ds_1 - B_2 \int_{-1}^1 \frac{(B_2)^{-1} K_2(s_2)}{s_2 - x_2} ds_2 = S$$

$$(10.11) \quad S = \pi(B_2 \{\text{Re}[P_2^+] + E_2\} - \frac{V_1}{V_2} \{\text{Re}[B_1^+ G_1^+ P_1^+] + \text{Re}[B_1^+ G_1^+] E_1\}) \\ + \frac{V_1}{V_2} \frac{1}{2} \{Q^{(1)} \zeta^{(1)} \delta(x_1 - z_1^{(1)}) + Q^{(2)} \zeta^{(2)} \delta(x_1 - z_1^{(2)})\}$$

Therefore, solving the integral equations for $K_2 = V_2$ on $[-b, b]$ and substituting the result into representations (10.3) and (10.4) yields solutions for W_1 and W_2 . These solutions satisfy all of the boundary conditions except when the incident wave reflects off a mixed interface. This is clear from integral equation (10.10) since the poles and delta functions on the right side cannot be

represented by the integrals on the left. In order to make the equation valid, the singularities on $[-b, -1]$, $[1, b]$ must be removed by suitable poles in P_2 , written as

$$(10.12) \quad P_2 = (p_2^{(1)} + i q_2^{(1)}) \frac{1}{z_2 - x_2^{(1)}} + (p_2^{(2)} + i q_2^{(2)}) \frac{1}{z_2 - x_2^{(2)}}$$

These P_2 poles are continuations of those due to reflected waves in (9.16), just as P_1 poles in (9.6) are continuations of those due to transmitted waves in (9.12).

Removal of the singularities requires an expression for the limit of poles on the real axes. According to (9.1) this is

$$(10.13) \quad \lim_{y \rightarrow 0^+} \frac{1}{z_n - x_n^{(i)}} = \frac{1}{x_n - x_n^{(i)}} \mp i \pi \delta(x_1 - x_1^{(i)})$$

i.e., the Cauchy integral of a delta function yields a simple pole by virtue of the sifting property. Evaluating P_1^+ and P_2^+ by taking the limit of poles in (9.6) and (10.12), and substituting into S gives two singular algebraic equations as $x_n \rightarrow z_n^{(i)}$, one in terms of poles and the other in terms of delta functions, namely

$$(10.14) \quad \pi B_2 p_2 \frac{1}{x_2 - x_2^{(i)}} \pm \frac{v_1}{v_2} \frac{(Q^{(i)})^2 \zeta_1^{(i)}}{1 + (Q^{(i)})^2} \frac{1}{x_1 - x_1^{(i)}} \rightarrow 0$$

$$\pi B_2 q_2 \delta(x_2 - x_2^{(i)}) - \frac{v_1}{v_2} Q^{(i)} \zeta_1^{(i)} \left\{ \frac{\pi}{1 + (Q^{(i)})^2} - \frac{1}{2} \right\} \delta(x_1 - x_1^{(i)}) \rightarrow 0$$

Solving the two by changing variables and taking the limit gives the complex coefficient in P_2 as

$$(10.15) \quad p_2^{(i)} + i q_2^{(i)} = \frac{v_1}{v_2} \frac{Q^{(i)} \zeta_1^{(i)}}{\pi B_2^{(i)}} \left(\frac{dx_2}{dx_1} \right)^{(i)} \left\{ \frac{i}{1 + i Q^{(i)}} - \frac{i}{2\pi} \right\}$$

where

$$(10.16) \quad \frac{dx_2}{dx_1} = \frac{v_2}{v_1} \frac{\omega_1}{\omega_2} \frac{1}{Q} \frac{x_2}{x_1} \frac{\sqrt{x_2^2 - 1}}{\sqrt{1 - x_1^2}}$$

11. Analysis of Undetermined Functions

The singular integral equations derived above involve multiplicative branch cut functions B_1 and B_2 , and additive entire functions E_1 and E_2 , which are still undetermined. They must of course be known completely before integral equations (10.8) and (10.10) can be solved. Their final form is found here by insuring that the integral representations exhibit the proper behavior at infinity and that the Cauchy-type integrals themselves exist.

First, consider the behavior of solutions at infinity in the

z_1 and z_2 planes. Through the characteristic mapping, these limit points correspond to bisecting points on the diffracted fronts, namely, B and E in Fig. 4. They are ordinary points in the sense that solutions are no more singular there than anywhere else on the diffracted fronts. It is readily established from (4.13) that solutions W_1 and W_2 are bounded on maps of the diffracted fronts except at isolated singular points where reflected or transmitted waves terminate. Consequently, representations (10.3) and (10.4) must be bounded at infinity, i.e., $O(1)$, in the z_1 and z_2 planes. Examining integral representations in the previous section, it is clear that as $z \rightarrow \infty$, P and the integrals themselves are $O(1/z)$, while from (9.8) $G_1 \rightarrow e^{O(1/z)} = 1$. Therefore, (10.3) and (10.4) yield the asymptotic estimate

$$(11.1) \quad W \approx B\{E + O(1/z)\} = O(1)$$

Since (9.9) and (9.14) imply that $B(z)$ is $O(z^d)$ at infinity for d an integer, and the above order condition requires BE to be $O(1)$, then entire function $E(z)$ is necessarily a polynomial, with real coefficients to make $E(x)$ real. Therefore, the only choices that satisfy condition (11.1) are—in order of increasing degree in $E(z)$ —either B is $O(z)$ and E is identically zero, or B is $O(1)$ and E is a constant, or B is $O(1/z^d)$ and E is a polynomial of degree d (positive).

Because all boundary conditions are satisfied by the integral representations for $W(z)$, and the behavior at infinity is correct, in principle, any of the above choices for $B(z)$ and $E(z)$ is admissible provided that the integrals exist. This is true because the mathematical diffraction problem is well-posed and must have a unique solution, to which any admissible choice of $B(z)$ and $E(z)$ should lead. The “proper” choice is defined operationally as the one that minimizes the number of constants that must be evaluated, namely, degree of polynomial $E(z)$. According to (9.9) and (9.14), B_1 and B_2 exhibit square root branch points at $\pm a$ and $\pm b$ respectively and have arbitrary integer order at infinity. Therefore, the simplest choice satisfying all of the above requirements is

$$(11.3) \quad B_1(z_1) = \sqrt{a^2 - z_1^2} \quad , \quad B_2(z_2) = \sqrt{b^2 - z_2^2} \quad , \quad E_1 = E_2 \equiv 0$$

assuming that the resulting integrals in (10.3) and (10.4) exist.

The question of integrability, i.e., existence of the integrals, is answered by examining solutions near the integration endpoints. Some indication of the endpoint behavior is provided by explicit representation (8.4) for W^\pm . Note that near $\rho = c_n$, $n = 1, 2$, the roots in (8.4) may be transformed using L'Hospital's rule as

$$(11.4) \quad \sqrt{c_n^2 - \rho^2} \approx \frac{c_n \omega_n}{\pi} \sqrt{1 - x_n^2}$$

Since (8.4) indicates that $V_2 = K_2$ vanishes like a square root at $x_2 = \pm 1$, B_1 may have its square root branch point in the numerator, thereby canceling K_2 's root in the first integral of (10.8) and (10.10). In order for the second integrals in (10.8) and (10.10) to exist at $\pm b$, B_2 may have its branch point in either the numerator or denominator. Thus, the simple choice given by (11.3) is admissible.

Finally, note that near the origin, (8.4) indicates that W^\pm is bounded when $v_1 \neq v_2$ (TE case), however, when $v_1 = v_2$ (TM case) the denominator becomes

$$(11.5) \quad v_1 c_2 \sqrt{c_1^2 - \rho^2} - v_2 c_1 \sqrt{c_2^2 - \rho^2} \approx \frac{c_1^2 - c_2^2}{2\mu_0 c_1 c_2} \rho^2$$

Although this appears to indicate that W^\pm is more singular in the TM case, from the discussion following asymptotic solution (4.11), R_1 and R_2 in (8.4) must vanish like ρ^2 so that solutions go to the same imaginary constant as $z \rightarrow 0$. Therefore, the Cauchy-type integral representations are valid near the origin, and moreover, $V_2 = K_2$ goes to a constant there.

DISCUSSION AND CONCLUSIONS

12. Discussion

This paper develops a theory of planar, vector wave diffraction for the generalized, two-wedge problem. The development proceeds from the general to the specific, namely, a general formulation of the initial-boundary problem in two complex half-planes, to a specific analysis for the case of $c_1 > c_2$, i.e., where the larger exterior wedge supports the slower wave. It is natural to call this case "thick wedge" diffraction, in contrast to "thin wedge" diffraction, where the thick ($>\pi$) or thin ($<\pi$) wedge is defined as the one supporting the slower wave. In either case the wave is always incident from the larger wedge. For example, consider Fig. 1 and imagine the slow wedge as glass and the fast wedge as air. Thick wedge diffraction, drawn on the left, involves a "concave" glass wedge ($>\pi$), so-called reentrant, with the wave incident from within the slower glass, whereas thin wedge diffraction drawn on the right involves a "convex" glass wedge ($<\pi$) with the wave incident from the faster air.

The reason for specializing analysis to the thick wedge is that multiple reflections do not occur, hence, it yields the simplest system of reflected, transmitted, and Mach waves. Consequently, the thick wedge represents the canonical problem—in the sense that it involves the simplest hyperbolic system while exhibiting all of the diffraction phenomenon's vector character. In contrast, for the thin wedge, i.e., $c_1 < c_2$, the transmitted waves may undergo multiple reflections, and more to the point, when the wedge angle is less than $2\cos^{-1}c_1/c_2$ the Mach waves overlap on the diffracted front, and for angles less than $\cos^{-1}c_1/c_2$ they eventually reflect between the interfaces. Mach wave overlap necessitates a more tedious algebraic elimination of hyperbolic solutions from boundary conditions on the real axes. This couples W_2 on multiple segments of the x_2 axis to W_1 on a semi-infinite segment of the x_1 axis. Therefore, the vector Riemann-Hilbert problem on W_1 is defined on a semi-infinite line, and so is the resulting singular integral equation. Although the additional algebra and manipulation required for the thin wedge are straightforward, such details are beyond the scope of the present paper.

The analysis presented here is, in most part, formal since rigorous mathematical justification does not accompany each step. Therefore, in order to claim that this is indeed a solution to the problem, extensive verification is necessary. To begin this process of verification, an evaluation of the singular integral equation was performed assuming the following parameters, $c_1 = 1.$, $c_2 = .6854$, $\omega_1 = 120^\circ$, $\omega_2 = 240^\circ$, $\phi = 60^\circ$ for both TM ($v_2/v_1 = 1$) and TE ($v_2/v_1 = .4697$) polarizations. These parameters correspond to the case of thick wedge diffraction pictured in Fig. 1c, for an exterior wedge of glass ($c_2 = .6854$) and interior wedge of air ($c_1 = 1.$). The actual numerics are described in detail elsewhere, see Wojcik and Mould in this report, so only a general description and results will be mentioned here.

The evaluation begins with a solution of singular integral equation (10.10). The standard approach to solving an integral equation is to first replace the integral by an approximate quadrature formula. The equation's unknown function is evaluated at discrete points distributed along the line of integration according to the type of quadrature formula used. By forcing satisfaction of the resulting approximate quadrature equation at an appropriate set of evaluation points on or off the quadrature points, the integral equation may be reduced to a determinate algebraic system of linear equations in terms of discrete values of the unknown function. The system can then be solved for the discrete unknowns using a standard linear system solver.

In contrast to a standard integral equation, the equation to be solved here is defined on two real lines, rather than just one, and exhibits two types of singularities, Cauchy-type principal values at $s_n = x_n$, $n = 1, 2$, and various square roots at the endpoints. Neither of these anomalies presents any fundamental difficulty since all of the known singularities may be subtracted and integrated exactly, leaving smooth integrands in the equation, while the integrals defined on separate lines are approximated by quadrature formulas appropriate to each line. The only concession is that the evaluation points cannot coincide with quadrature points or endpoints, because of the singularities.

It is most efficient to use an N-point Gauss quadrature formula over each "logical" line segment, i.e., corresponding to mixed and nonmixed segments on the two interfaces, namely, $(-a, 0)$, $(0, a)$ on x_1 and $(-b, -1)$, $(-1, 0)$, $(0, 1)$, $(1, b)$ on x_2 . On this basis, the smoothed unknown is approximated by four $2N-1$ degree polynomials over the four distinct segments. Note that there are four rather than six polynomials because of the one-to-one mapping between segments $(\pm a, 0)$ and $(\pm 1, 0)$. The polynomial coefficients become the approximate quadrature equation's unknowns. Evaluation points in sufficient numbers are arbitrarily interspersed between quadrature points to render a determinate algebraic system on the coefficients. This system, consisting of perhaps 128 unknowns in the case of ten quadrature points per segment, is readily solved using direct Gaussian elimination. The right hand side of the system is proportional to the incident wave only, hence, the homogeneous algebraic system corresponds to vanishing input.

The integral equation has been solved in this way for the problem parameters mentioned above and no difficulties were encountered. Substitution of the results back into the scalar and vector Riemann-Hilbert problems shows that all are satisfied to the numerical accuracy of the integral equation's solution. This solution yields the unknown densities, K_1 and K_2 , in integral representations for $W_1(z_1)$ and $W_2(z_2)$ respectively. Thus, the solution at any point in either plane is found by performing the quadrature in (10.3) or (10.4). It is natural to use the same Gauss quadrature formula for this evaluation as that used for the integral equations themselves since the polynomial solutions are then integrated exactly.

With $F'(w) = W(z(w))$ known in the elliptic subdomains, it is necessary to integrate F' to obtain F , hence $f(\rho, \theta)$. In the hyperbolic domains this is simply done by integrating plane wave solutions (7.7) by inspection to yield step functions for the incident, reflected, and transmitted waves. In the elliptic subdomains, f is obtained by numerically integrating f_ρ in the second of (4.9) along a radial path starting from known initial conditions at the diffracted front, then computing all other points in the domain by integrating f_θ in the first of (4.9) along circumferential paths emanating from the initial radial path. The explicit singularities in the solution, including those of C_+ in (4.9), must be accommodated by subtracting them, integrating the singular terms exactly, and numerically integrating the smooth residual. Since f_ρ has poles where the reflected and transmitted waves are tangent to the diffracted front, these points are best approached in the θ direction. Finally, solutions in Mach wave regions are computed by continuing along characteristics from the mixed boundary where it is known from the diffracted wave calculation in the interior wedge. These integrations were verified in the elliptic domains by differencing the solutions and confirming that they satisfy the boundary conditions and Cauchy-Riemann conditions to the numerical accuracy of the calculation.

A final verification of the solution was performed by comparing to a finite element simulation of the diffraction problem. Since discrete numerical solutions like finite elements cannot capture step response due to grid dispersion, a wavelet exhibiting minimal dispersion was input to the grid. The exact solution was then convolved with this wavelet for comparison at a number of grid points around the vertex. The agreement was excellent, with any discrepancies attributable to the finite element model.

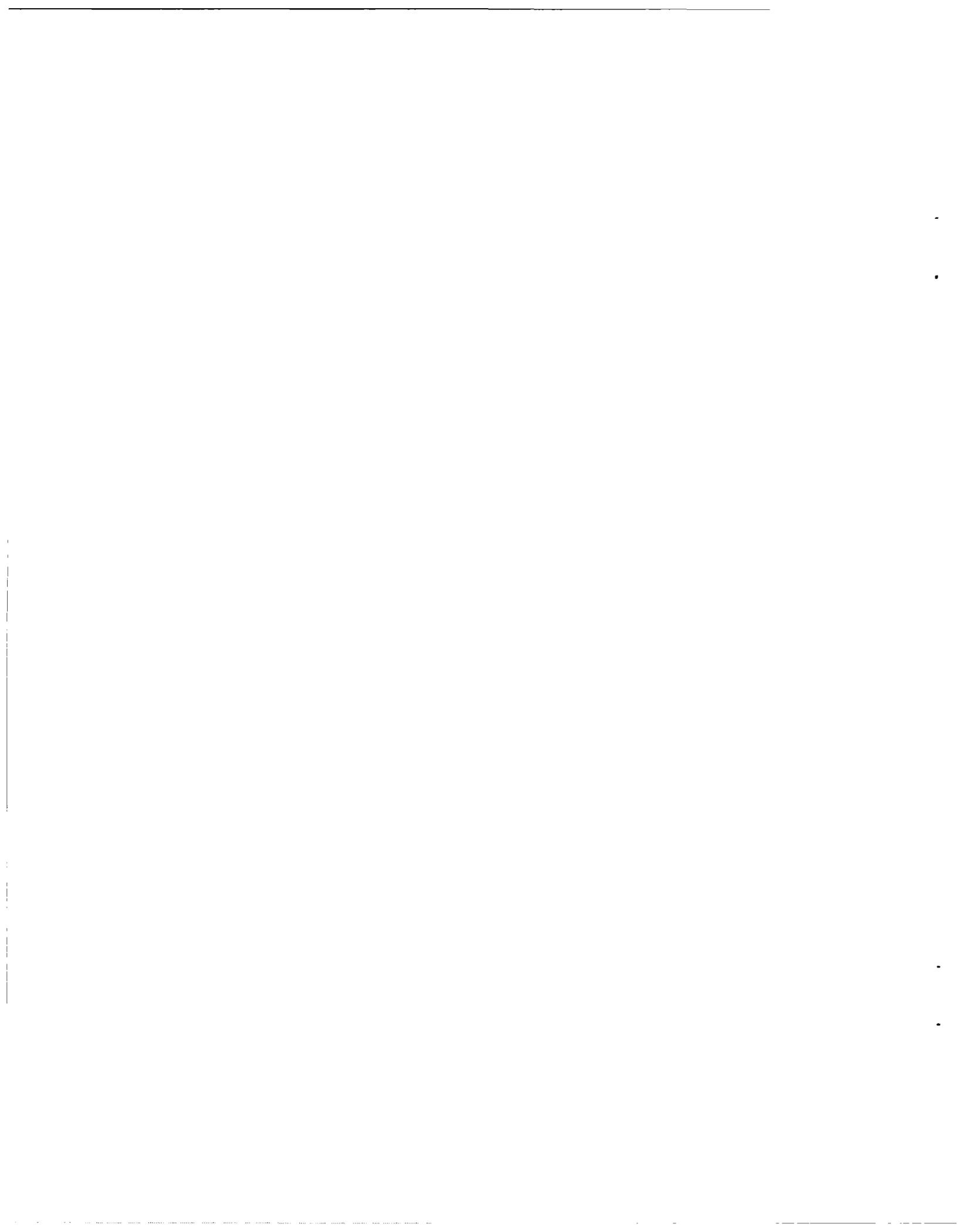
13. Conclusions

In seeking solutions to initial-boundary problems, the function of applied mathematical analysis is to reduce complicated, unsolved problems to simpler, "solved" problems. In this vein, the preceding analysis has reduced the canonical generalized vector diffraction problem to a well-posed singular integral equation on finite segments of two real axes. Solving this integral equation yields the density functions in integral representations of the unknown analytic functions, from which a complete solution of the original problem follows by transformation and quadrature.

This is certainly an effective reduction of the diffraction problem since numerical solution techniques for singular integral equations are well-known and very effective. The need to ultimately solve integral equations should not come as a surprise since most scalar diffraction problems are naturally posed as integral equations, e.g., see Carrier, et al. (1966) in reference to dual integral equations. Apparently, a closed-form solution to a vector problem of this complexity is beyond any reasonable expectation. In any event, this paper is concluded by claiming that the canonical problem of light diffraction by planar dielectric wedges is now essentially solved.

REFERENCES

- [1] Carrier, G.F., M. Krook, and C.E. Pearson (1966). Functions of a Complex Variable, McGraw-Hill Book Company, New York.
- [2] Harrington, R. F. (1961). Time-Harmonic Electromagnetic Fields, McGraw-Hill Book Company, New York.
- [3] Keller, J. B. and A. Blank (1951). 'Diffraction and reflection of pulses by wedges and corners,' *Comm. Pure Appl. Math.*, **4**.
- [4] MacDonald, H. M. (1902). Electric Waves, Cambridge University Press.
- [5] Muskhelishvili, N. I. (1958). Singular Integral Equations, Noordhoff International Publishing, Leyden.
- [6] Pauli, W. (1938). 'On asymptotic series for functions in the theory of diffraction of light,' *Phys. Rev.*, **54**, pp. 924-931.
- [7] Sommerfeld, A. (1896). 'Mathematische theorie der diffraction,' *Math. Ann.*, **47**, pp. 317-374.
- [8] Sommerfeld, A. (1964). Partial Differential Equations in Physics, Academic Press, New York.
- [9] Stratton, J. A. (1941). Electromagnetic Theory, McGraw-Hill Book Company, New York.
- [10] Wojcik, G. L. (1989). 'A Numerical and Theoretical Study of Seismic Wave Diffraction in Complex Geologic Structure,' Final Report AFGL-TR-88-0291, Air Force Geophysics Laboratory Contract F19628-84-C-0102, Weidlinger Associates.
- [11] Wojcik, G. L. (1989). 'Vector Boundary Problems in Connected Complex Half-Planes,' included in [10] above.
- [12] Wojcik, G. L. & J. Mould, Jr. (1989). 'Light Diffraction by a Planar Air-Glass Wedge: Verification of the Vector Theory,' included in [10] above.



LIGHT DIFFRACTION BY A PLANAR AIR-GLASS WEDGE: VERIFICATION OF THE VECTOR THEORY[†]

by

Gregory L. Wojcik & John Mould, Jr.
Weidlinger Associates*

ABSTRACT

The diffraction of light by dielectric wedges occurs whenever electromagnetic waves illuminate a sharp edge on an interface separating different dielectric media, e.g., air and glass. Such edges are often encountered in practice, but despite their common occurrence and intrinsic importance, the associated diffraction phenomenon has never been quantified due to the lack of an effective mathematical theory.

This paper presents the numerical evaluation of a solution based on a new, mathematical vector diffraction theory for the planar canonical problem of two, contiguous, semi-infinite, dielectric wedges, with a TM or TE polarized plane wave incident from the larger wedge and parallel to the edge. A detailed solution and results are presented for a symmetric example that minimizes complexity while exhibiting all of the vector character.

This solution is verified against a finite element simulation of the same problem and excellent agreement is found. Consequently, the theory provides a benchmark against which to validate discrete numerical simulations of models with sharp edges that diffract vector waves.

[†] Supported under AFGL Contract F19628-84-C-0102 and NSF SBIR Grant SI-8760089

* 4410 El Camino Real, Suite 110, Los Altos, CA 94022

1. Introduction

This paper evaluates a canonical problem in the theory of light diffraction by dielectric wedges. The physical problem consists of an exterior 240° glass wedge "enclosing" an interior 120° air wedge. This two-wedge diffractor is illuminated by a plane electromagnetic wave incident from within the glass and parallel to its edge. Transverse electric (TE) and transverse magnetic (TM) plane wave polarizations are considered. Problem geometry is illustrated in Fig. 1, showing a two-dimensional slice perpendicular to the edge. By virtue of wedge geometry and wave incidence angle, the problem is two-dimensional. For convenience, only the symmetric problem will be calculated, where the wave is perpendicular to the wedge's bisector.

The purpose of the paper is to verify a mathematical theory of vector wave diffraction recently developed by Wojcik [2]. This theory takes advantage of the self-similarity exhibited by wedge-type diffraction problems lacking any characteristic length scale. It is assumed that the incident plane wave has a simple step in field amplitude across the wave front. More general inputs, e.g., time-harmonic, follow by convolution with the step solution's first time-derivative. For the relative simplicity of working with the same order of derivative in boundary conditions, relations on the field's normal and tangential gradients are used, rather than equivalent relations on the field and its normal gradient. This choice yields a formulation in terms of field gradient unknowns, requiring a simple numerical quadrature on the solution to recover the field itself.

Self-similarity reduces the governing two-dimensional wave equation in each wedge to a one-dimensional "string" equation in mixed characteristic coordinates. The term, mixed, means that the equation changes type over different parts of the domain, namely, elliptic inside the diffracted wave fronts and hyperbolic outside. A Schwarz-Christoffel transformation maps the characteristic domains to two complex half-planes. The problem's initial and boundary conditions all map to the real axes of these two half-planes and yield either elliptic, hyperbolic, or mixed conditions there.

Thus, by means of self-similarity and domain mapping, the diffraction problem is reduced to the determination of two analytic functions, each defined in its own complex plane. This is the so-called normalized boundary problem. Boundary conditions relate real and imaginary parts of the two functions on the real axes. These boundary relations effectively define a vector generalization of the Riemann-Hilbert problem, where analytic functions are defined by a linear relation between their real and imaginary parts. A vector theory for this class of boundary problems is given in [3].

A solution of the normalized boundary problem begins by expressing the analytic functions as a Cauchy-type integral representation in each plane. These representations consist of Cauchy integrals on segments of the real axes and include certain additive and multiplicative algebraic factors. Substituting into the pure elliptic boundary conditions yields a singular integral equation in one unknown but on the two real lines. The additive and multiplicative factors are determined from the hyperbolic and mixed boundary conditions.

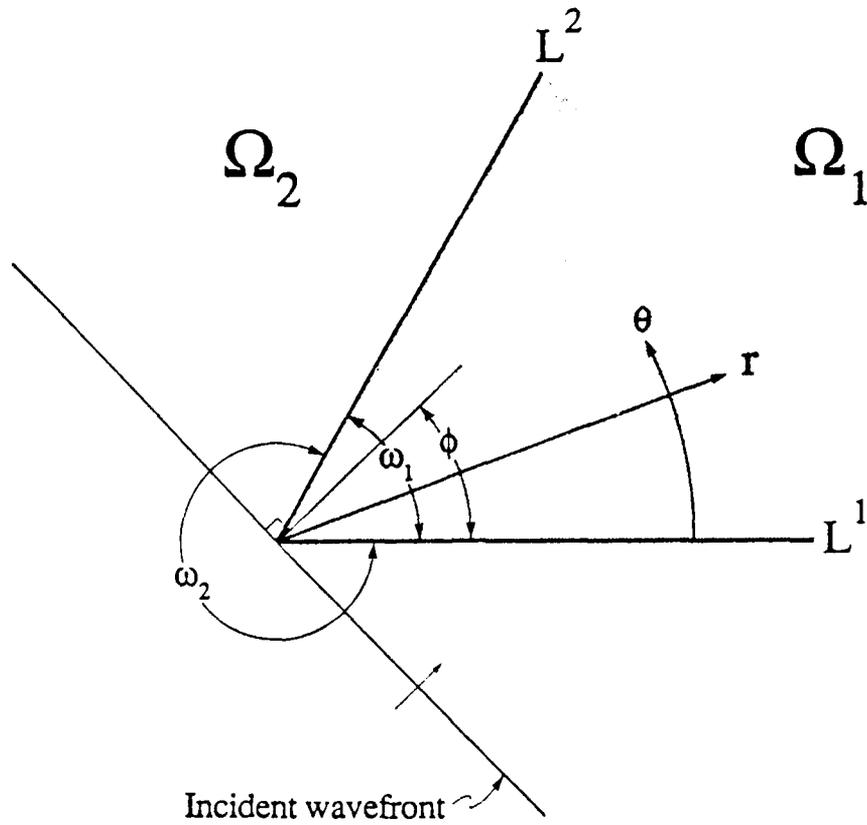


Figure 1. Geometry and polar coordinates for the two-dimensional, two-wedge diffractor. Ω_1 and Ω_2 denote the interior and exterior wedges, respectively

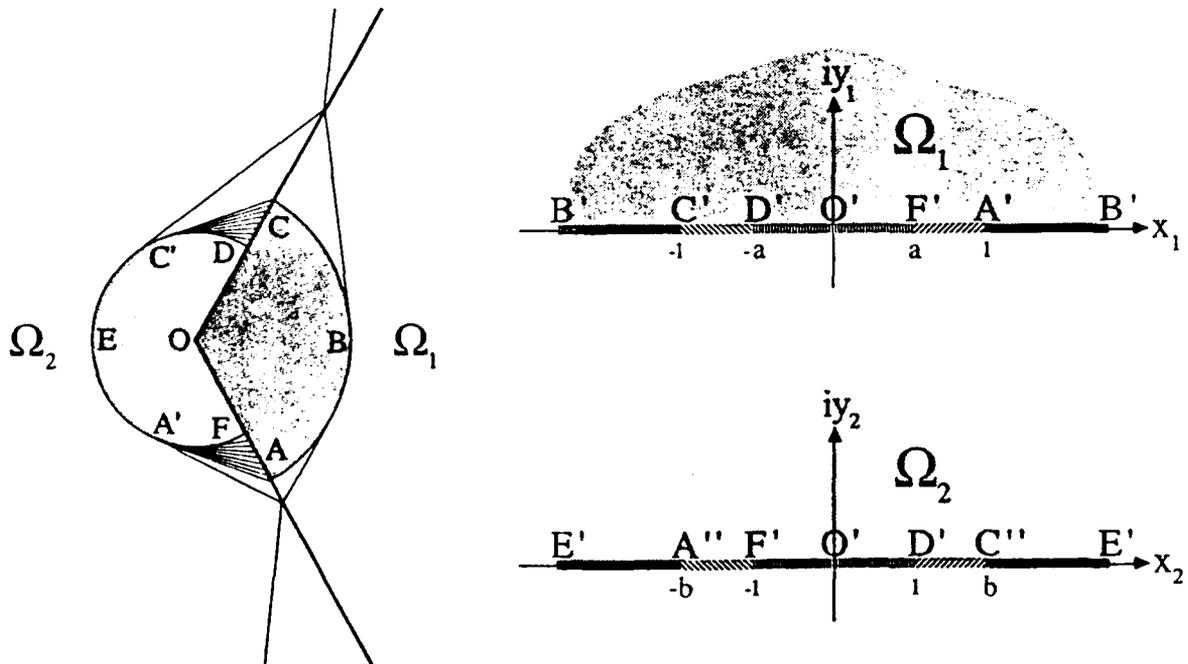


Figure 2. The system of incident, reflected, transmitted, and diffracted wave fronts is drawn on the left. On the right is the mapping of the two wave domains to complex half-planes.

Therefore, the canonical diffraction problem is converted to that of solving a singular integral equation. Given a solution to this equation, the boundary integral representations are evaluated to provide spatial gradients of the diffracted electromagnetic field at each point in the physical two-wedge diffractor. The field itself is recovered by numerical integration of the gradients, resulting in the complete diffracted field for a step function input. If another input time-function is of interest, typically time-harmonic, then this last quadrature is unnecessary because convolution is performed on the derivative of the step function response, which is given directly by the boundary integral representations.

2. The Singular Integral Equation

The problem considered in this paper is chosen because it yields the simplest system of plane waves associated with the diffracted waves, while exhibiting all of the phenomenon's vector character. The system of wave fronts is drawn in Fig. 2, showing incident, reflected, and transmitted plane waves, circular diffracted waves, and "triangular" Mach wave zones excited in the slower wedge by superluminal grazing incidence of the faster diffracted wave. Note that all plane waves are tangent to their diffracted front. By way of contrast, if the interior and exterior wedges are glass and air, instead of vice versa, then the transmitted plane waves can undergo multiple reflections and the Mach waves can intersect on the diffracted front, significantly complicating the analysis with no better elucidation of the solution process.

Also shown in Fig. 2 is the mapping of the domains to complex half-planes. Observe the one-to-one mapping of segments C'D'O'F'A' and C''D'O'F'A'' on the real axes back to the air-glass interface. The unknown analytic functions defined in the complex half-planes are denoted $W_n = U_n + iV_n$, $n = 1, 2$, and are represented by the boundary integrals

$$(2.1) \quad W_n(z_n) = B_n(z_n) \left\{ \frac{1}{\pi} \int_{-p_n}^{p_n} \frac{B_n^{-1} V_n}{s_n - z_n} ds_n + P_n(z_n) \right\}$$

where functions B_n , P_n , and interval $[-p_n, p_n]$ are determined from the boundary conditions. On intervals $-1 < x_1 < 1$ and $-b < x_2 < b$ of the half-plane boundaries, the analysis in [1] yields a singular integral equation to be solved for the imaginary part of the analytic function in the Ω_2 -plane, namely,

$$(2.2) \quad \text{Re}[B_1] \int_{-a}^a \frac{(B_1^+ Q)^{-1} V_2}{s_1 - x_1} ds_1 - B_2 \int_{-b}^b \frac{B_2^{-1} V_2}{s_2 - x_2} ds_2 = S$$

where

$$(2.3) \quad S \equiv \pi \text{Re}[B_2^+ P_2^+ - \frac{v_1}{v_2} B_1^+ P_1^+]$$

is a smooth source term proportional to the incident wave's amplitude through transmission and

reflection coefficients in P_1 and P_2 respectively, e.g., see [1]. Constants v_1 and v_2 depend on the incident wave's polarization: in the TM case $v_1 = v_2 = 1/\mu_0$ where μ_0 is free-space magnetic permeability; in the TE case $v_i = 1/\epsilon_i$, $i = 1, 2$ where ϵ_i is the dielectric permittivity. The remaining functions in (2.2) are

$$(2.4) \quad B_1 = \sqrt{a^2 - z_1^2} e^{\Gamma_1(z_1)} \quad , \quad B_2 = \sqrt{b^2 - z_1^2}$$

$$(2.5) \quad \Gamma_1(z_1) = \frac{2}{\pi} \int_a^1 \frac{s_1 \tan^{-1} Q}{s_1^2 - z_1^2} ds_1$$

$$(2.6) \quad Q = \frac{c_1 v_2}{c_2 v_1} \frac{\sinh\left\{\frac{\omega_2}{\pi} \cosh^{-1} \frac{1}{|x_2|}\right\}}{\sinh\left\{\frac{\omega_1}{\pi} \cosh^{-1} \frac{1}{|x_1|}\right\}}$$

3. Solution of the Integral Equation

Solutions of singular integral equation (2.2) are calculated assuming the following parameters, $c_1 = 1.$, $c_2 = .6854$, $\omega_1 = 120^\circ$, $\omega_2 = 240^\circ$, and $\phi = 60^\circ$, for both TM ($v_2/v_1 = 1$) and TE ($v_2/v_1 = .4697$) polarizations. These correspond to symmetric incidence of a plane wave from an exterior wedge of glass ($c_2 = .6854$) onto an interior wedge of air ($c_1 = 1.$).

The standard approach to solving an integral equation is to first replace the integral by an approximate quadrature formula. The equation's unknown function is sampled at discrete points distributed along the line of integration according to the type of quadrature formula used. By evaluating the resulting approximate quadrature equation at an appropriate set of evaluation points on or off the quadrature points, the integral equation may be reduced to a determinate algebraic system of linear equations in terms of discrete values of the unknown function. The system can then be solved for the discrete unknowns using a standard linear system solver. In principle, any desired degree of accuracy can be achieved by choosing a sufficiently large number of quadrature points.

In contrast to a standard integral equation, the present case is defined on two real lines, rather than just one, and exhibits two types of singularities, Cauchy-type principal values at $s_n = x_n$, $n = 1, 2$, and various square roots at the endpoints. Neither of these anomalies presents any fundamental difficulty. All of the known singularities may be subtracted and integrated exactly, leaving smooth integrands to be integrated numerically. The resulting nonsingular version of (2.2) is in terms of unknown residual $R_2(x_2)$ —obtained by dividing out the square root singularities in V_2 . To accommodate integrals defined on separate lines, they are simply approximated by quadrature formulas appropriate to each line. The only concession is that evaluation points cannot

coincide with quadrature points or endpoints, because of the singularities.

The unknown piecewise smooth function R_2 is represented by separate polynomials over each "logical" integration interval. These subintervals corresponding to mixed and nonmixed segments on the two interfaces in Fig. 2, namely, $(-a,0)$, $(0,a)$ on x_1 and $(-b,-1)$, $(-1,0)$, $(0,1)$, $(1,b)$ on x_2 . In each subinterval, N_{eval} interior evaluation points or nodes are chosen and four $N_{eval}-1$ degree Lagrangian polynomials are postulated to represent the smooth unknown in terms of nodal values over the four subintervals. There are four rather than six polynomials because of the one-to-one relation between segments $(\pm a,0)$ and $(\pm 1,0)$.

For the numerical integration we choose a Gauss-Legendre quadrature rule using $N_{quad} = N_{eval}/2$ quadrature points, which integrates the polynomial representation of the residual exactly. Thus, integral equation (2.2) is written once for each of the N_{eval} nodes on x_2 in each of the four intervals. Each of these equations also references the residual value at the N_{quad} quadrature points of each interval, so Lagrangian interpolation equations are written to express the residual at these points in terms of its values at the evaluation nodes. Additionally, N_{quad} quadrature points are chosen in the two x_1 -plane subintervals in order to approximate the nonsingular version of the first integral in (2.2). These discrete equations constitute an $N \times N$ linear system (where $N \equiv 4 \times N_{eval} + 6 \times N_{quad} + 8$) that is solved for the unknown nodal values of R_2 .

This linear system, consisting of perhaps 128 unknowns in the case of ten quadrature points per segment, is readily solved using direct Gaussian elimination. Note that the right hand side of the system is proportional to the incident wave, hence, the homogeneous algebraic system corresponds to vanishing input.

The integral equation has been solved for the problem parameters mentioned above and no numerical difficulties were encountered. A comparison of residual solution R_2 using four-, six-, and ten-point quadrature rules is shown in the upper plot of Fig. 3. This indicates that a ten-point rule per subinterval is near the limit of observable error. Multiplying residual R_2 by the appropriate square root singularities yields the imaginary parts, $V_1 = (v_1/v_2)QV_2$, in integral representations (2.1) for $W_1(z_1)$ and $W_2(z_2)$. The real part of this representation on the real axis yields U_1 and U_2 compared in the lower plot of Fig. 3. Plots of real and imaginary parts on the integration intervals in each plane are shown in Fig. 4. Substitution of the results back into the vector Riemann-Hilbert problems, i.e., the boundary conditions, shows that all are satisfied to the numerical accuracy of the integral equation's solution. For example, the dashed curve showing $v_1 U_1$ in the lower plot of Fig. 3 is proportional to the gradient normal to the wedge interface. This quantity should be continuous across the interface, and indeed, the solid curve shows that the error, normalized by the peak value of $v_1 U_1$ and scaled on the right axis, is negligible. It appears to be proportional to the slope of $v_1 U_1$ and, as expected, is greatest at the endpoints where the quadrature solutions are extrapolated rather than interpolated.

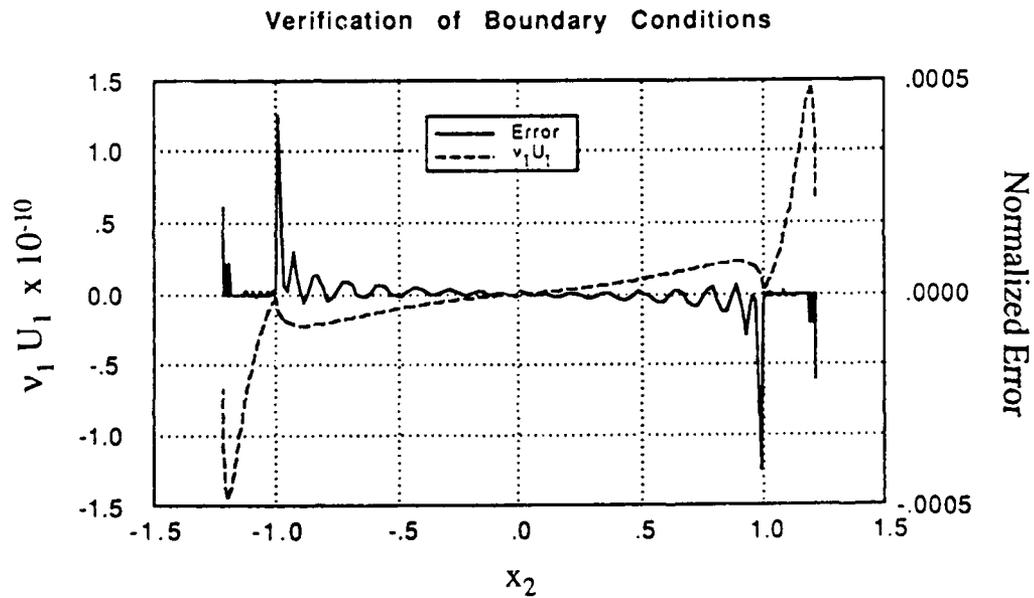
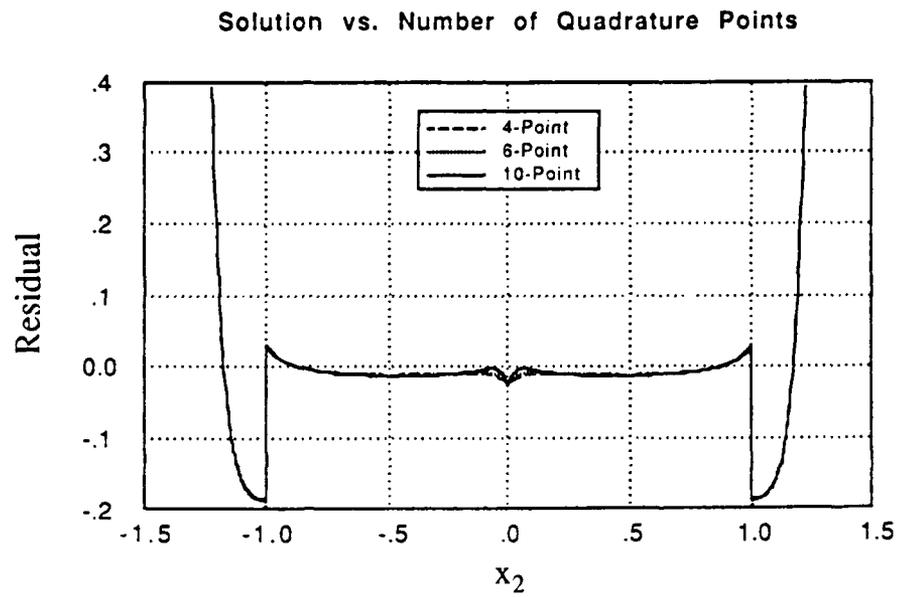


Figure 3. Convergence of residual solution with quadrature level (upper), and error in solution of the integral equation for the ten-point quadrature (lower).

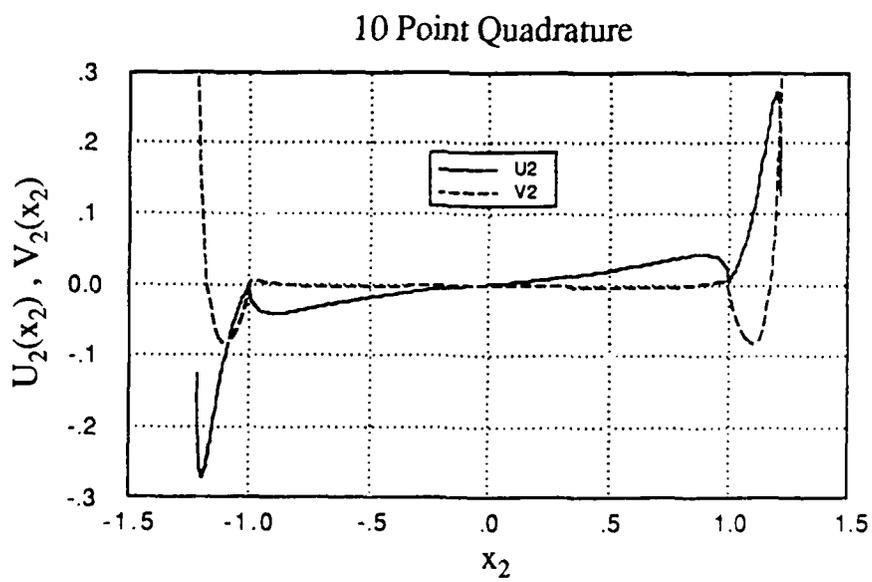
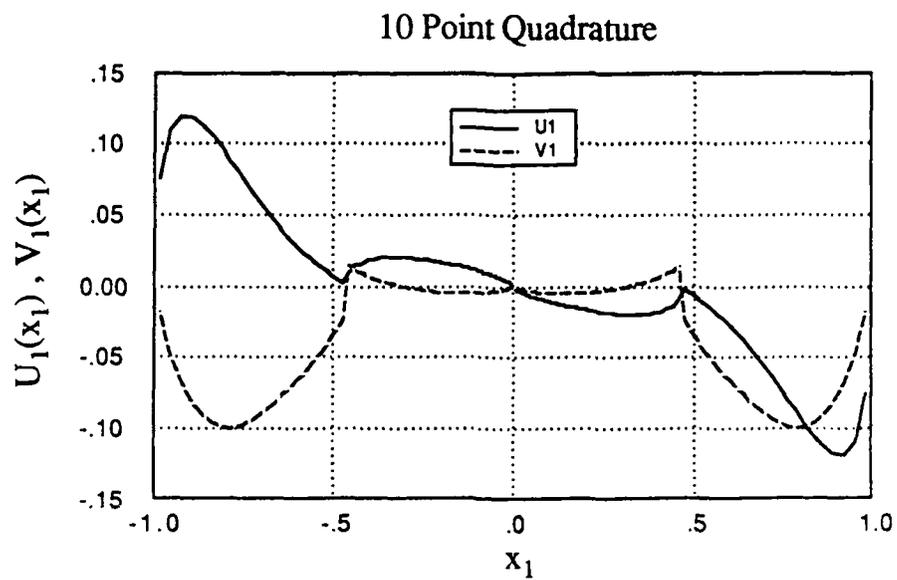


Figure 4. Calculated real (U_n) and imaginary (V_n) parts of W_n on the real axes.

It is noted that alternate, and perhaps more efficient techniques could be employed to solve singular integral equation. For instance, one could construct a weighted quadrature scheme incorporating the endpoint singularities so that they need not be subtracted out. For this application however, such an approach would require a different quadrature scheme for each interval. At the other extreme, one could think of fitting a high order (our approach retains only the constant term) polynomial to the kernel and extending the subtraction process until the numerical quadrature becomes unnecessary. The drawback here is that the analytical integrals such become increasingly more complicated and result in more involved bookkeeping to assemble the coefficient matrix.

4. Evaluation of Diffracted Field Gradients

With the imaginary parts known, the solution gradients at any point in either plane is found by performing the integration in (2.1). It is natural to use the same Gauss-Legendre quadrature formula for this evaluation as was used for the integral equations themselves because the polynomial solutions for the residuals are then integrated exactly. Special care must again be taken to subtract out all square root endpoint singularities and integrate them exactly. This should also be done for the Cauchy singularity since the representation is often evaluated quite close to the real axis, where the singularity overwhelms the quadrature scheme.

The physical solution to be determined is $f(\rho, \theta)$ where $\rho \equiv r/t$ and f is equal to electric field E_z for the TM case and magnetic field H_z for the TE case. Both H_z and E_z field components are parallel to the edge of the diffractor. The analysis in [1] gives this solution as

$$(4.1) \quad f_n(\rho, \theta) = F_{n+}(w_{n+}) + F_{n-}(w_{n-}) \quad , \quad w_{n\pm} \equiv \theta \pm i \cosh^{-1} c_n / \rho$$

where only the derivative with respect to $w_{n\pm}$ is determined from the boundary conditions. This derivative is in fact $W_n(z_n(w_{n\pm}))$ determined above in the half-planes corresponding to the elliptic subdomains.

5. Evaluation of Diffracted Fields

It is necessary to integrate $F_n' = W_n$ to obtain F_n , hence $f_n(\rho, \theta)$. In the hyperbolic domains this is simply done by integrating plane wave solutions by inspection to yield step functions for the incident, reflected, and transmitted waves. In the elliptic subdomains, f_n is most conveniently obtained by numerically integrating $(f_n)_\rho$ along a radial path starting from known initial conditions at the diffracted front, computing all other points within the diffracted zone by integrating $(f_n)_\theta$ along circumferential paths emanating from the initial radial path. The explicit singularities in the solution must be accommodated by subtracting them, integrating the singular terms exactly, and numerically integrating the smooth residual. Since $(f_n)_\rho$ has poles where the reflected and transmitted waves are tangent to the diffracted front, these points are best approached in the θ

direction. Finally, the solutions in Mach wave regions are computed by continuing along characteristics from the mixed boundary where it is known from the air wedge diffracted wave calculation.

The solution was integrated on a dense grid of points covering the diffracted wave region and contour plots of the results are shown in Fig. 5. These illustrate and contrast the general character of TM and TE diffracted fields for an incident step wave. One-dimensional plots along the interface and wedge bisector are drawn in Fig. 6, indicating the stronger diffracted fields generally excited by TM waves. These integrations were verified in the elliptic domains by differencing the solutions and confirming that they satisfy the boundary conditions and Cauchy-Riemann conditions to the numerical accuracy of the calculation.

6. Finite Element Model Comparisons

A final verification of the solution was performed by comparing to a finite element simulation of the diffraction problem. Since discrete numerical solutions like finite elements cannot capture step response due to grid dispersion caused by space and time discretization, a wavelet exhibiting minimal dispersion was input to the grid. The exact solution described above was then convolved with this wavelet for comparison at a number of grid points around the vertex. The finite element model, Cartesian (stairstep) discretization, and the output point locations are depicted in Fig. 7.

Comparisons are shown in Fig's. 8 and 9 for TE and TM polarizations, respectively. For the two cases agreement is quite reasonable, within about one percent of the incident wave amplitude in the interior wedge and similarly in the exterior. Agreement appears poorer in the exterior wedge only because the diffracted signal is much weaker there. The comparison is good even near the vertex. Note that the precursor signal obvious in the exterior wedge is the incident plane wave, which is sufficiently separated in time to allow a reasonably clean diffraction.

Discrepancies in the TM case are attributable to the finite element model, since this stairstep model does not even include a sharp vertex. Despite minor discrepancies, it appears that the finite element simulation generally verifies the exact solution.

7. Conclusions

The overall goal of this paper is to confirm solvability of the integral equations resulting from a new mathematical theory of vector wave diffraction given in [1], and verify the resulting solutions to the extent feasible. This goal has been achieved and results support the conclusion that the theory does indeed solve this historically intractable problem. It is also gratifying to note that the theory yields an integral representation of solutions that is readily evaluated to a high degree of precision without "heroic" efforts, either numerically or computationally.

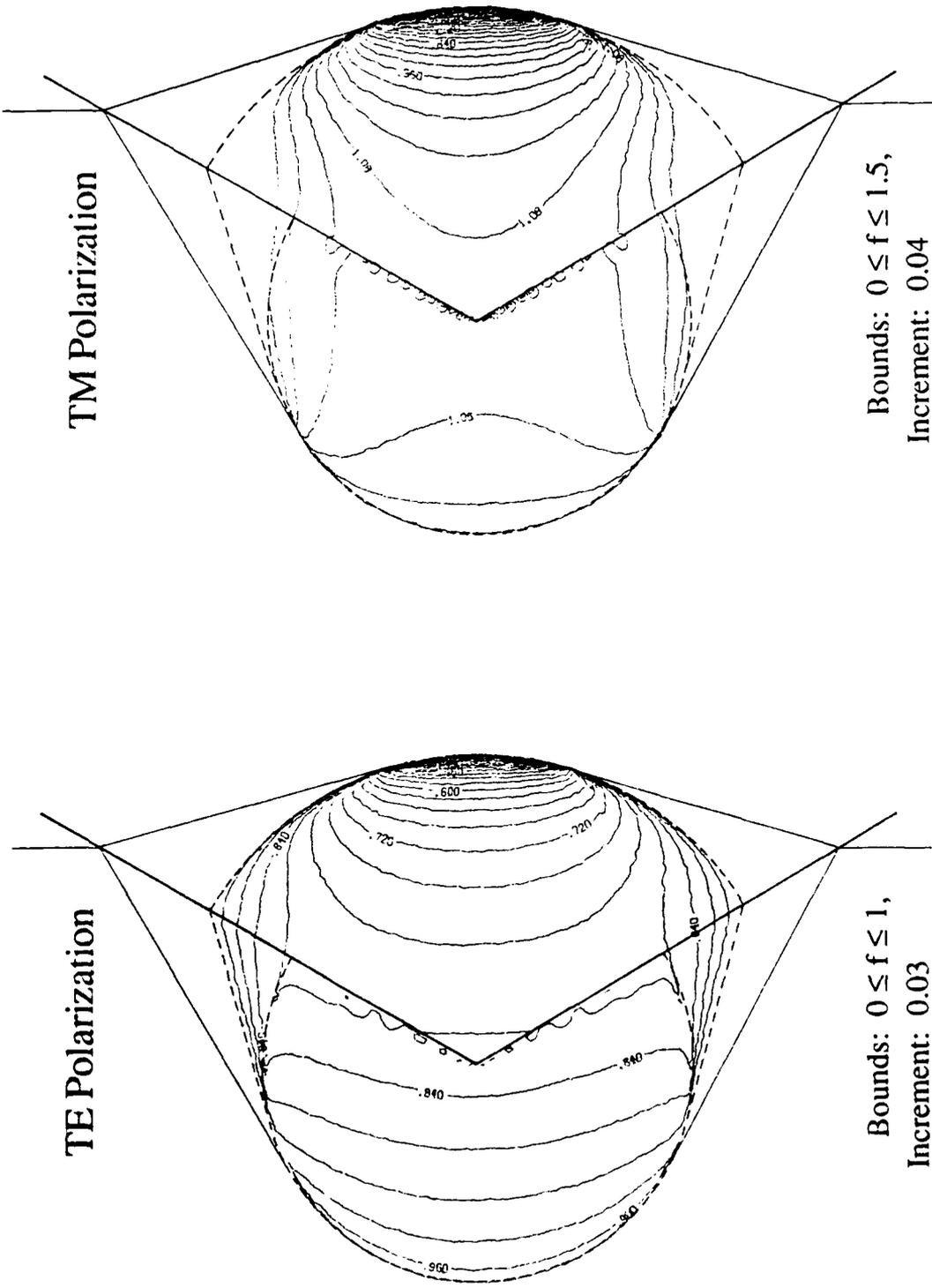


Figure 5. Contour plots of the full diffraction solution for TE and TM polarizations.

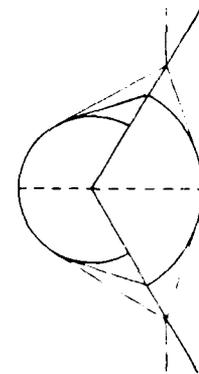
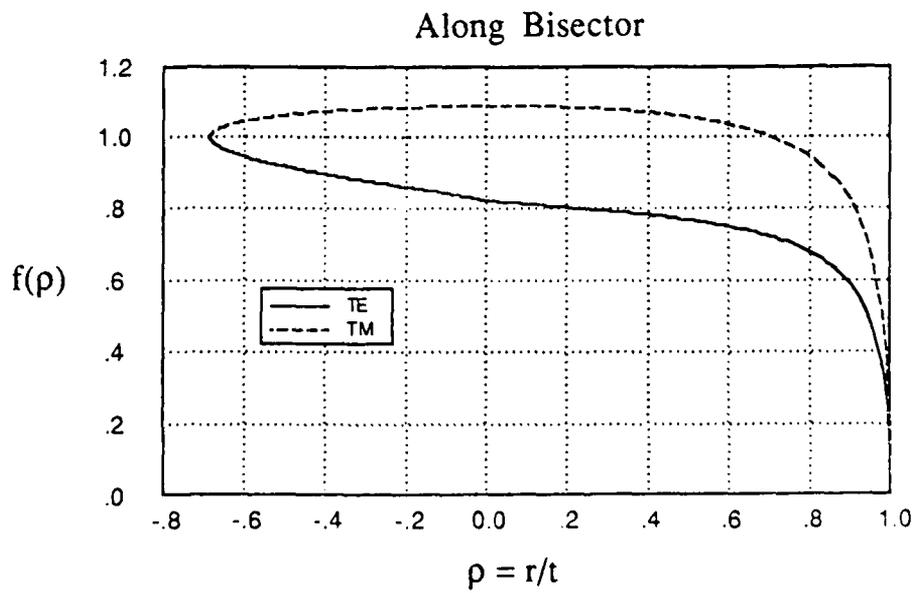
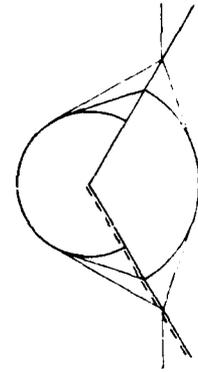
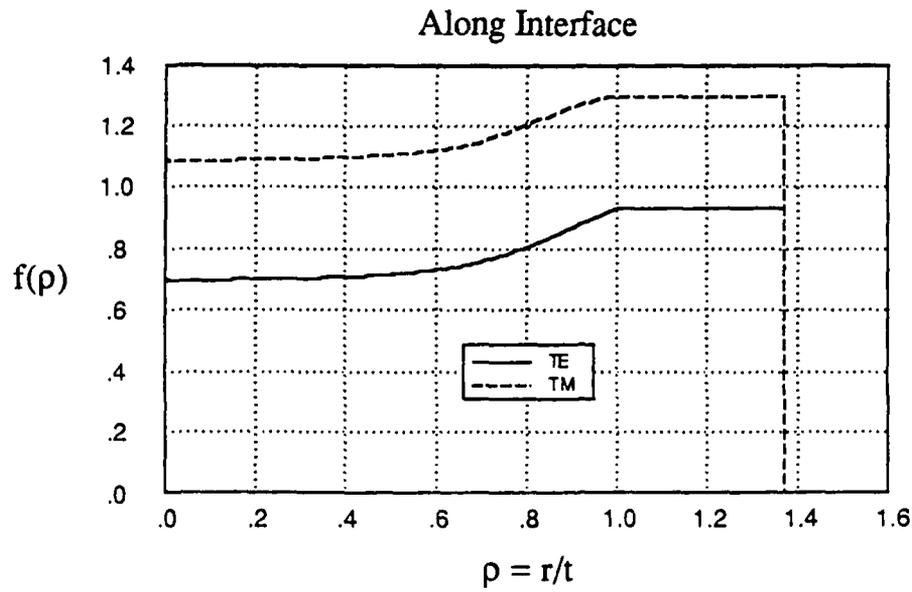


Figure 6. Comparison of TE and TM analytical solutions on the wedge interface (upper) and on the bisector (lower).

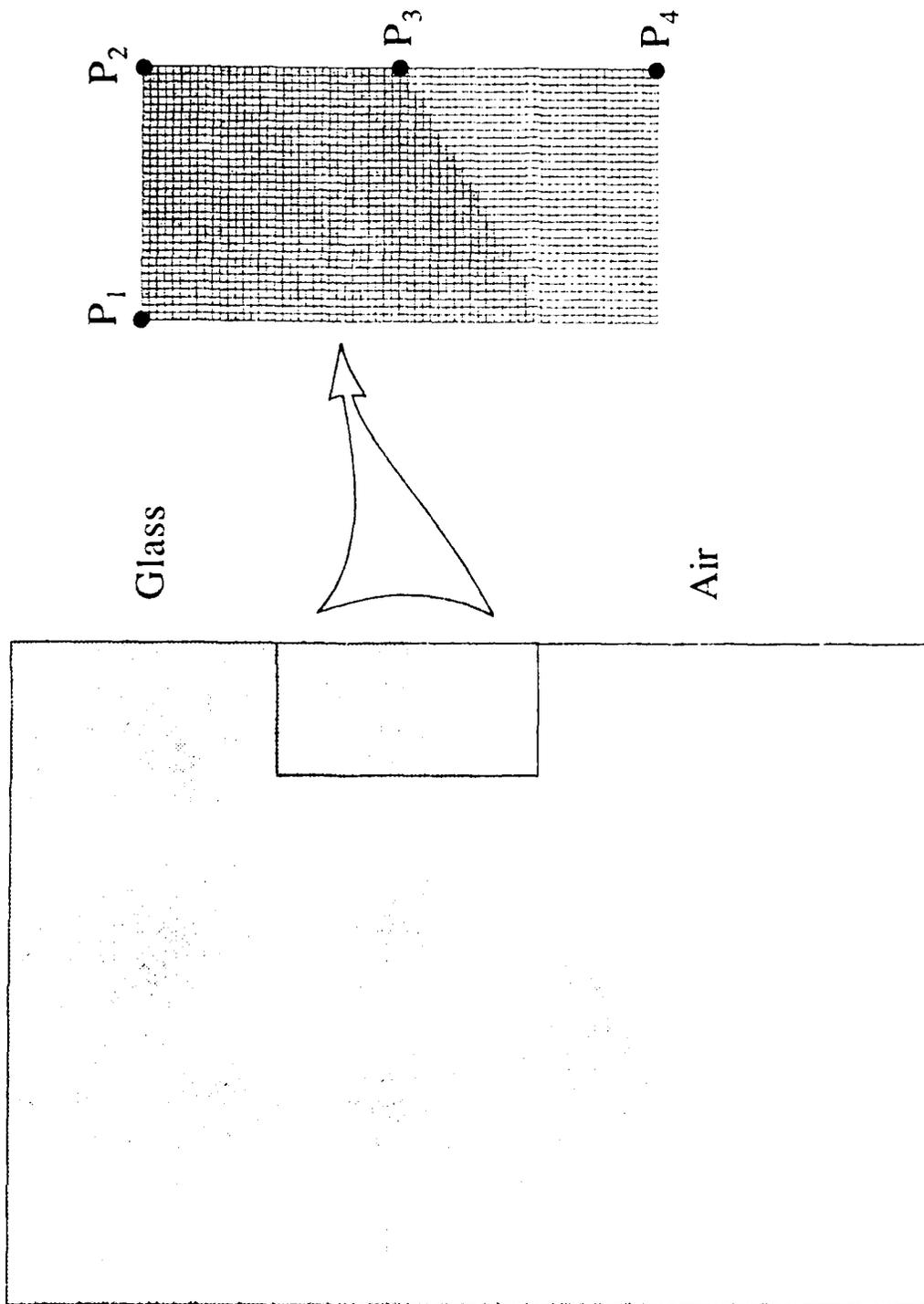


Figure 7. The finite element model used for verifying the analytical solution, showing staircase discretization of the interface and the output point locations.

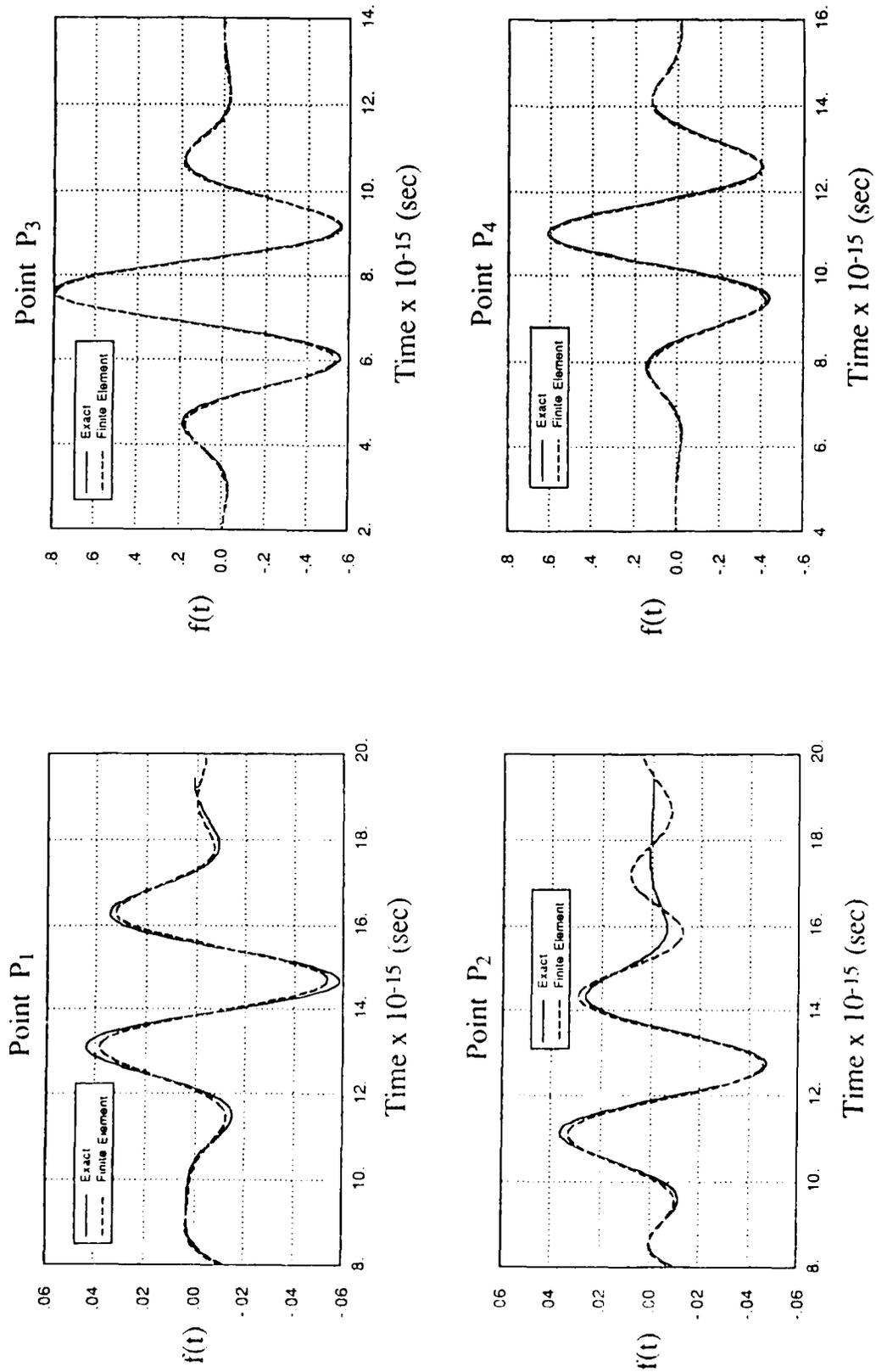


Figure 8. Comparisons of finite element and analytical diffraction solutions for TE polarization. The incident wave is outside the time window at P₁ and P₂.

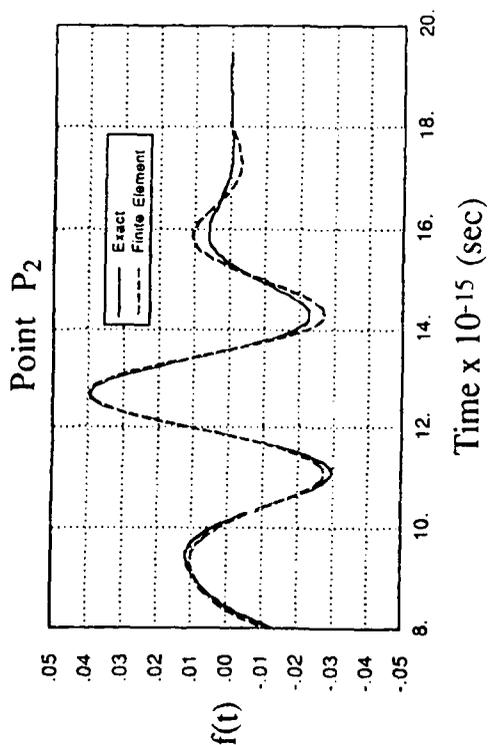
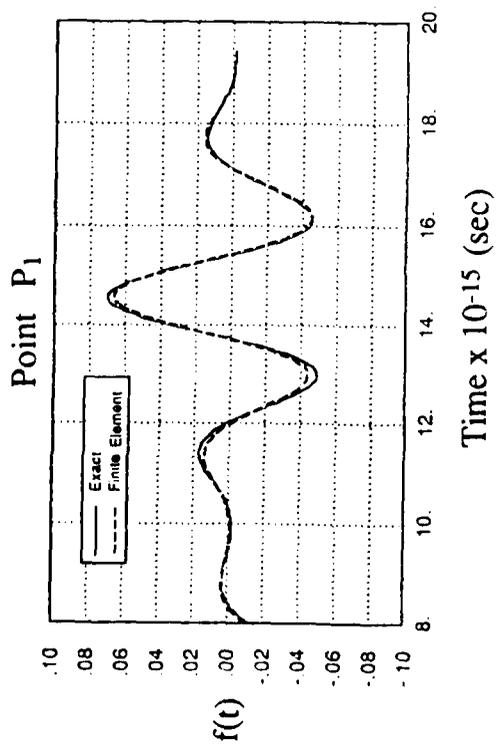
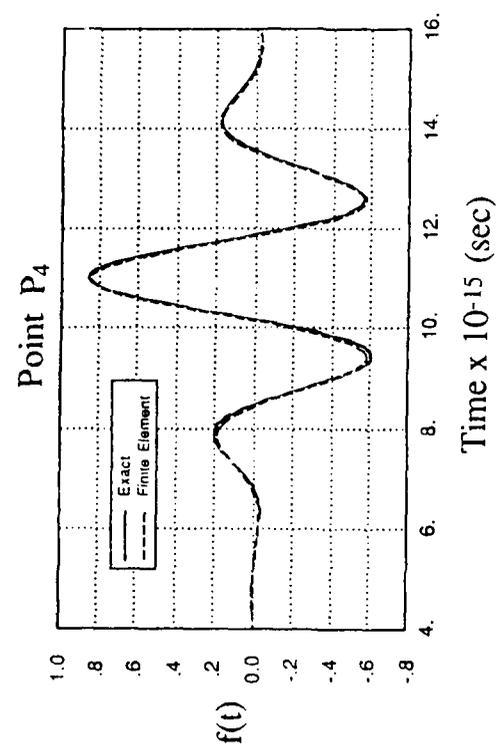
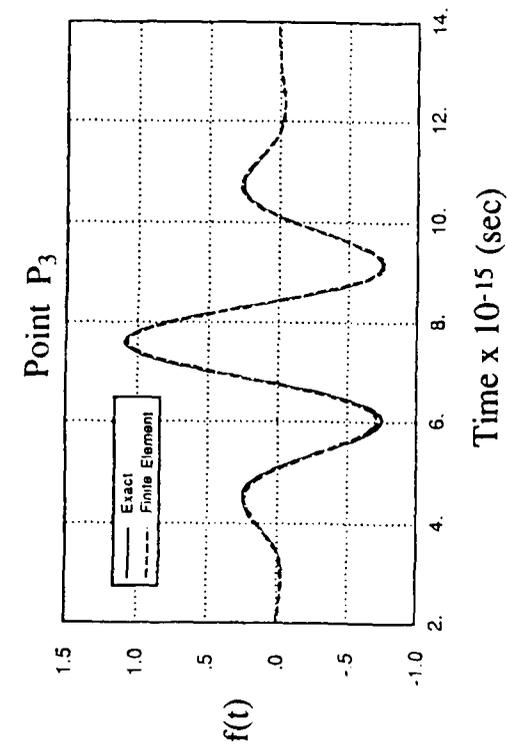


Figure 9. Comparisons of finite element and analytical diffraction solutions for TM polarization. The incident wave is outside the time window at P₁ and P₂

On this basis it seems reasonable to continue the verification exercise in the future with other geometries and incidence angles, particularly in conjunction with finite element simulations. However, the simulations should examine effects of conforming versus stairstep interface approximations, to see if this minimizes the discrepancy near the vertex. An observation relevant to finite element modeling is the apparent breakdown in accuracy of these diffraction solutions as an edge singularity becomes more pronounced. For other diffraction problems, particularly in elasticity where the singularity is known to be algebraic, stronger edge singularities may invalidate certain discrete wave propagation models that include sharp edges and corners. The solution reported here provides the first benchmark that can be used as a rigorous basis to test the effectiveness of finite element, finite difference, and boundary element methods in modeling vector wave diffraction in ideally "rough" media.

REFERENCES

- [1] Wojcik, G. L. (1989). 'A Numerical and Theoretical Study of Seismic Wave Diffraction in Complex Geologic Structure,' Final Report AFGL-TR-88-0291, Air Force Geophysics Laboratory Contract F19628-84-C0102, Weidlinger Associates.
- [2] Wojcik, G. L. (1989). 'A Mathematical Theory of Light Diffraction by Dielectric Wedges,' included in [1] above.
- [3] Wojcik, G. L. (1989). 'Vector Boundary Problems in Connected Complex Half-Planes,' included in [1] above.

VECTOR BOUNDARY PROBLEMS IN CONNECTED COMPLEX HALF-PLANES[†]

by

Gregory L. Wojcik
Weidlinger Associates*

ABSTRACT

Harmonic boundary problems in connected, plane domains present formidable analytical difficulties. In principle, they may be solved by iteration, e.g., Schwarz's alternating procedure. This paper presents a more elegant alternative in the form of vectorized integral representations for plane, multiply connected problems. The theory provides a methodology for solving some previously intractable multi-domain boundary problems in physics and mechanics.

The problem is to determine the N -vector of analytic functions, $W(z) = U(x,y) + iV(x,y)$ given the linear relation, $aU - bV = G$ on limit points of the connected domains. These domains are mapped conformally to N complex half-planes, each supporting one component of $W(z)$, and boundary conditions are prescribed on all or parts of the real axes. Complex variable $z = x + iy$ is a diagonal $N \times N$ matrix with each element spanning one of the N planes, while boundary coefficients a and b are full $N \times N$ matrices and G is an N vector of distributions.

An analytical vector basis is derived in terms of so-called Schwarz dual boundary equations and their normal forms. These are solved in terms of a vector of Cauchy-type line integrals and certain homogeneous solutions. The boundary conditions are converted to conjugate vector integral equations over segments of the N real axes using a Hilbert-type transform pair. Either conjugate equation must be solved numerically for the kernel of the corresponding integral representation since general closed-form solutions do not appear feasible.

[†] Supported under Air Force Geophysics Laboratory Contract F19628-84-C-010

* 4410 El Camino Real, Suite 110, Los Altos, CA 94022

1. Introduction

A fundamental boundary problem in the theory of complex variables is to find scalar function $W(z) = U(x,y) + iV(x,y)$ analytic on its domain of definition, given linear relation $aU - bV = G$ (i.e., $\text{Re}\{(a + ib)W\} = G$) between continuous real and imaginary parts on the domain's limit points. The problem is associated with the names of Riemann (from his famous thesis (1851), e.g., [7]) and Hilbert [3] for their seminal contributions to the case of a closed contour. Carleman's name should be appended for his closely related work on singular integral equations [2]. The practical scope has since been broadened by analysts in various disciplines to systems of contours and curved lines, but typically in only one complex variable (domain or medium), e.g., see Muskhelishvili [6], and later references in the applied mathematics and mechanics literatures.

The objective of this paper is to extend the analytical theory of two-dimensional boundary problems from one to several complex variables—in other words, from simply to multiply connected domains. The similar case of multiply connected harmonic domains in two and three dimensions was formally solved by H.A. Schwarz using an alternating domain procedure (iteration) on scalar functions, and also by Poincaré as the balayage method, e.g., see [2] and [4]. In contrast to these formalisms, the present analysis is accomplished by vectorization of the complex variable basis and conformal mapping of physical domains, so that each unknown function is supported in its own complex half-plane—essentially one element in a "vector" of half-planes. General vector solutions are found as Schwarz-type integral representations on the union of boundaries, with the density vectors determined by solving well-posed systems of Cauchy-type singular integral equations. Closed-form boundary integral expressions are only available under certain degenerate conditions on matrix boundary functions a and b , of which the scalar theory (single complex variable) provides the simplest example.

Motivation for the paper is a need to solve classical partial differential equations with nonseparable boundary conditions in multiply connected domains. These problems arise whenever one or more domains support two or more harmonic functions coupled nontrivially on common boundaries. The theory's kernel is provided by the problem of self-similar vector wave diffraction—where several wave functions are coupled on a plane boundary or interface with discontinuous tangent [8]. The theory supplies a practical basis for other difficult multi-function or multi-domain problems. These may involve inverse formulations as well as the direct approach described here.

Considering the variety of boundary shapes encountered in practice, the following analysis only examines the canonical problem defined in so-called parallel complex half-planes, with boundary conditions given on the real axes. Here the term parallel simply means that the domains meet at their limit points and are otherwise disconnected, like parallel lines. Isolated singularities may be prescribed off the real axes, while the behavior at infinity (and perhaps special boundary

points) is specified in each half-plane for completeness. Connected polygonal domains reduce to this canonical case via Schwarz-Christoffel transformations. More general domains can in principle be mapped conformally to circles by virtue of Riemann's mapping theorem, and then to half-planes. The mapping may in addition be multiple-valued; however, the original boundaries must have single-valued parameterizations.

The paper is organized into eleven sections. An example supplying background and motivation is described in Sec. 2. A vectorized form of the general mathematical problem is formulated in Sec. 3. Since this form is not amenable to the classical (scalar) Riemann-Hilbert approach, an analytical vector basis in terms of so-called Schwarz dual boundary equations and their normal forms is derived in Sec. 4. General solutions of the normal form are developed in Sec. 5 and the dual boundary equations are solved in Sec. 6, yielding conjugate Schwarz integral representations and the Hilbert-type transform pair. The boundary conditions are converted to conjugate integral equations over segments of the real axes in Sec. 7 using the Hilbert-type transforms. Section 8 returns to the classical Riemann-Hilbert form and examines a degenerate closed-form solution that does not involve solving integral equations. In Sec. 9 the prescribed isolated singularities and endpoint conditions are incorporated in the representations via the homogeneous part of general solutions. Overall results are discussed in Sec. 10 and the paper is concluded in Sec. 11.

2. Background

This section formulates a class of plane, multiply connected boundary problems in terms of the vector equations analyzed in the paper. This class involves two unknown harmonic functions and illustrates the lowest dimension vector (versus scalar) formulation. The results provides background and motivation for the analytical development in subsequent sections, although they do not yield the most general expression covered by the theory.

For all two-component problems, the canonical (mapped) domains consist of two complex half-planes, designated Ω_n and spanned by complex variables $w_n = u_n + iv_n$, $n=1,2$. Each half-plane supports a single harmonic field denoted by unknown function $X_n(u_n, v_n)$. Segments of the real axes across which the fields communicate are designated L_1 and L_2 in the half-planes. These segments, finite or infinite, are maps of the common boundary connecting the two original domains, hence there is a one-to-one correspondence between points on L_1 and L_2 .

In conventional applications to equilibrium problems in connected domains, the boundary coupling consists of two continuity conditions—one on the fields and the other on their normal derivatives. These are natural generalizations of Dirichlet and Neumann conditions for disjoint domains. For example, in heat conduction or electrostatics the two conditions would represent continuity of temperature and heat flux, or electrostatic potential and electric field intensity.

respectively. In other problems the conditions are not so simply interpreted, but general expressions can be defined on the fields and gradients.

Since a mixture of conditions involving the fields and their derivatives is awkward analytically, it is convenient to replace field continuity by continuity of its tangential derivative. Thus, both boundary conditions are expressed uniformly in terms of gradients, either normal or tangential to the interfaces. Note that recovery of the unknown harmonic functions using this formulation requires an additional quadrature after solving the resulting boundary problem for dX_n/dz_n . However, this is preferable to working throughout with mixed boundary conditions, i.e., involving the functions and their derivatives.

The normal- and tangential-derivative boundary conditions can usually be written in one of two more general forms on segments L_1 and L_2 . For two unknown functions coupled on the boundary these conditions are either

$$(2.1) \quad \begin{array}{l} a_1 \frac{\partial X_1}{\partial u_1} + a_2 \frac{\partial X_2}{\partial u_2} = G^1 \\ b_1 \frac{\partial X_1}{\partial v_1} + b_2 \frac{\partial X_2}{\partial v_2} = G^2 \end{array} \quad \text{or} \quad \begin{array}{l} a_1 \frac{\partial X_1}{\partial u_1} + b_2 \frac{\partial X_2}{\partial v_2} = G^1 \\ b_1 \frac{\partial X_1}{\partial v_1} + a_2 \frac{\partial X_2}{\partial u_2} = G^2 \end{array}$$

where coefficients a_n and b_n , $n=1,2$ incorporate the original boundary condition factors and the transformation of normal (v_n) and tangential (u_n) derivatives, while G^1 and G^2 are generalized functions representing boundary distributions (sources). The continuity conditions described above for electrostatics or heat conduction are a special case of the pair on the left. The pair on the right arises whenever normal and tangential derivatives are related. Both conditions occur in a variety of practical problems. Observe that each equation in (2.1) relates functions supported in separate half-planes, across segments L_1 and L_2 of the real lines. These functions are defined on the boundaries in terms of the original interface parameters, which typically reduce to an arc length on each distinct segment of the coupled interface.

The vector boundary problem for this example, as well as more involved cases, is formulated by defining complex functions

$$(2.2) \quad Z_n(w_n) = X_n(u_n, v_n) + i Y_n(u_n, v_n)$$

with $n=1,2$, where X_n is the unknown harmonic function defined above and Y_n is its conjugate. Since the derivative of Z_n on the real axes involves the gradients appearing in the boundary conditions, defining the new functions

$$\begin{aligned}
 W_n(z_n) &\equiv Z_n'(w_n) \\
 (2.3) \quad &= \frac{\partial X_n}{\partial u_n} - i \frac{\partial X_n}{\partial v_n} \\
 &= U_n(x_n, y_n) + i V_n(x_n, y_n)
 \end{aligned}$$

yields the components necessary for the algebraic vector formulation. Boundary conditions (2.1) are now given by linear equations on the U_n and/or V_n , written in matrix form as

$$(2.4) \quad \begin{bmatrix} a_1 & a_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} G^1 \\ G^2 \end{bmatrix}$$

or

$$(2.5) \quad \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} - \begin{bmatrix} 0 & b_2 \\ b_1 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} G^1 \\ G^2 \end{bmatrix}$$

Observe that these conditions yield disjoint coefficients, i.e., the matrices have no elements in common. Equations (2.4,5) define a particular case of the general vector boundary conditions studied in the following, written as

$$(2.6) \quad a U - b V = G$$

where a and b are full matrix coefficients, and U , V , and G are vectors. In practice, coefficients a and b may be discontinuous and it is convenient to restate (2.6) as

$$(2.7) \quad \operatorname{Re}(cW) = G, \quad c \equiv a + ib$$

where $W = U + iV$ is the unknown analytic vector. In this form there is no need to consider matrices a and b separately, but instead in the more natural combination, $c = a + ib$. Equations (2.6) or (2.7) yield the vector form of the boundary problem analyzed below.

3. The Canonical Half-Plane Problem

To state the general mathematical problem consider N complex variables, i.e., Argand diagrams,

$$(3.1) \quad z_n = x_n + i y_n \quad n = 1, N$$

and N unknown complex functions

$$(3.2) \quad W_n(z_n) = U_n(x_n, y_n) + i V_n(x_n, y_n) \quad n = 1, N$$

defined on $y_n > 0$, where U_n and V_n are real functions satisfying Laplace's equation, i.e., harmonic. Given a linear relation between the real and imaginary parts of all W_n coupled on lines L_n over each of the N real axes, namely, boundary conditions in the form

$$(3.3) \quad \sum_{m=1}^N [a_{mn} U_m(x_m, 0) - b_{mn} V_m(x_m, 0)] = G^n(x_n) \quad n = 1, N$$

the problem is to find the N functions, $W_n(z_n)$, analytic in the upper half-planes except at prescribed isolated singularities, and with prescribed endpoint behavior on L_n .

Each of the N equations in (3.3), in terms of given real boundary functions a_{mn} , b_{mn} , and distribution G^n , is defined generally on L_n . Each line L_n is a mapping to real axis x_n , which may be multi-valued (overlap on x_n) but is assumed to have a single-valued parameterization. The isolated singularities in each plane are specified by singularity function $S_n(z_n)$. This is any analytic representation regular on L_n with the same isolated singularities as $W_n(z_n)$ in the upper half-plane, including the point at infinity. When appropriate (a priori), the behavior of $W_n(x_n)$ at irregular points on L_n (endpoints and certain interior points) is specified by endpoint function $E_n(x_n)$, which merely exhibits the proper order or continuity conditions. An example is the order at infinity (if not included in S_n) which depends in part on the mapping of the original domain to the half-plane.

The boundary problem defined by (3.1,2,3) is better stated in vector form as

$$(3.4) \quad z = x + i y$$

$$(3.5) \quad W(z) = U(x,y) + i V(x,y)$$

$$(3.6) \quad a U_0 - b V_0 = \text{Re}\{(a + ib)W_0\} = G \quad \text{on } L$$

respectively, where L is the union of mappings and

$$(3.7) \quad W_0 = U_0 + i V_0 \equiv U(x, 0) + i V(x, 0)$$

Complex variable z and its real and imaginary parts x and y are $N \times N$ diagonal matrices, while complex function $W(z)$ and its parts U and V are N component vectors. Boundary condition coefficients $a \equiv [a_{mn}]$ and $b \equiv [b_{mn}]$ are full $N \times N$ matrices, and inhomogeneous term $G(x)$ is an N vector of distributions. The prescribed isolated singularities are given by analytic N vector $S(z)$ and the explicit endpoint behavior is exhibited in N vector $E(x)$. Observe that the above vector form is algebraically identical to the scalar form, and except for some necessary operational ordering no distinction need be made in the following analysis.

4. Dual Boundary Equations and Normal Forms

To express the vector boundary problem in a form amenable to analysis, vector unknown $W(z)$ is continued into the lower half-plane by (Schwarz) reflection about the real axis as

$$(4.1) \quad W(z^*) = W^*(z)$$

where superscript $*$ denotes a complex conjugate. It is assumed that all subsequent analytic functions introduced in this analysis are likewise defined in the lower half-plane by reflection, hence, will satisfy the reflection principle.

The real and imaginary parts of $W(z)$ can be written as the sum and difference of the function and its complex conjugate as

$$(4.2) \quad \begin{aligned} U &= \operatorname{Re}W = \frac{1}{2}(W + W^*) \\ V &= \operatorname{Im}W = \frac{1}{2i}(W - W^*) \end{aligned}$$

On the real line, replacing $W(x)$ and $W^*(x)$ by limits from above and below the axis, namely

$$(4.3) \quad \begin{aligned} W_+(x) &\equiv W(x) \\ W_-(x) &\equiv W^*(x) \end{aligned}$$

by virtue of reflection into the lower half-plane, and rewriting (4.2) yields the underlying Schwarz boundary equations governing analytic continuation off L ,

$$(4.4) \quad \begin{aligned} W_+ + W_- &= 2U_0 \\ W_+ - W_- &= i2V_0 \end{aligned}$$

In the following sections these conjugate equations are solved for $W(z)$ in terms of real or imaginary boundary values, yielding two Schwarz-type analytic representations. Either representation can in principal be used as a basis for solving the vector boundary problem.

Dual boundary equations (4.4) are most tractable after reduction to so-called Schwarz normal forms. The second of (4.4) is the vector prototype. More general forms are derived by introducing homogeneous matrix and vector equations corresponding to (4.4). Diagonal matrix functions $f(z)$ and $g(z)$ are defined, satisfying the homogeneous equations,

$$(4.5) \quad \begin{aligned} f_+ + f_- &= 0 \\ g_+ - g_- &= 0 \end{aligned}$$

Clearly, f is imaginary and g is real on L , and it is assumed that no terms vanish identically. Multiplying the first of (4.4) by f_+ , the second by g_+ , and replacing f_+W_- by $-f_-W_-$ and g_+W_- by g_-W_- yields the forms

$$(4.6) \quad \begin{aligned} f_+W_+ - f_-W_- &= 2f_+U_0 \\ g_+W_+ - g_-W_- &= i2gV_0 \end{aligned}$$

Defining vector functions $A(z)$ and $B(z)$ satisfying the homogeneous normal forms,

$$(4.7) \quad \begin{aligned} A_+ - A_- &= 0 \\ B_+ - B_- &= 0 \end{aligned}$$

whence A and B are real on L , and adding zero to both sides of (4.6) provides the most general normal forms as

$$(4.8) \quad \begin{aligned} (f_+ W_+ - A_+) - (f_- W_- - A_-) &= 2 f_+ U_0 \\ (g_+ W_+ - B_+) - (g_- W_- - B_-) &= i 2 g V_0 \end{aligned}$$

To complete this reduction it is necessary to normalize the homogeneous matrix boundary equations in (4.5) by introducing the so-called logarithmic normal form. The second equation is already normalized, but to uniformly convert both to a logarithmic form the pair is rewritten as

$$(4.9) \quad \begin{aligned} f_+ f_-^{-1} &= f_-^{-1} f_+ = -I \\ g_+ g_-^{-1} &= g_-^{-1} g_+ = I \end{aligned}$$

where I is the $N \times N$ identity matrix and the inverses exist by definition. Since the terms commute, taking the multiple-valued matrix logarithm gives the desired forms as

$$(4.10) \quad \begin{aligned} \ln f_+ - \ln f_- &= i \pi I + i 2 \pi j \\ \ln g_+ - \ln g_- &= i 2 \pi k \end{aligned}$$

where j and k are diagonal matrices of arbitrary integers. Thus, the original Schwarz dual boundary equations in (4.4) are equivalent to the two general normal forms in (4.8), with f and g defined by the logarithmic normal forms in (4.10).

5. Solutions of the Normal Forms

The normal boundary problems equivalent to dual equations (4.4) all have the vector or matrix form

$$(5.1) \quad F_+ - F_- = D(x)$$

on L , where $F(z)$ is a composition of analytic functions in the z plane with branch cut on L , over which distribution $D(x)$ is defined. Since $F(z)$ is defined in the lower half-plane via Schwarz reflection, $D(x)$ is just twice the imaginary part of F on L . Defining particular solution $C(z)$ such that

$$(5.2) \quad C_+ - C_- = D(x)$$

i.e., any factorization of $D(x)$, and subtracting (5.2) from (5.1) gives the homogeneous problem

$$(5.3) \quad (F_+ - C_+) - (F_- - C_-) = 0$$

satisfying

$$(5.4) \quad H_+ - H_- = 0$$

where $H_{\pm} \equiv F_{\pm} - C_{\pm}$ is continuous across L , i.e., $H_+ = H_- = H(x)$. Continuing off L by inspection (replacing x by z) yields the general solution as

$$(5.5) \quad F(z) = C(z) + H(z)$$

Thus $F(z)$ is the sum of particular and homogeneous solutions satisfying (5.2) and (5.3) respectively.

A particular solution follows by representing density $D(x)$ as an integral over L on the real axis. This is developed here using the sifting property of the delta function,

$$(5.6) \quad D(x) = \int_L ds \delta(s-x) D(s)$$

by factoring the diagonal matrix of real delta functions as the difference between upper and lower limits at a simple pole on L

$$(5.7) \quad \delta(s-x) = \frac{1}{2\pi i} ([s-x]_+^{-1} - [s-x]_-^{-1})$$

In other words, a simple pole on the real axis is the Green's function for the normal form. Replacing $\delta(s-x)$ in (5.6) and interchanging the order of integral and limit operators gives

$$(5.8) \quad D(x) = C_+[x; D] - C_-[x; D]$$

for which $C_{\pm}[x; D]$ are limiting values of the following Cauchy-type integral

$$(5.9) \quad C[z; D] = \frac{1}{2\pi i} \int_L ds [s-z]^{-1} D(s)$$

where s and z are diagonal matrices and D may be a vector or diagonal matrix. Indenting the path of integration around x on L gives these limits as

$$(5.10) \quad C_{\pm}[x; D] = \pm \frac{D(x)}{2} + C_p[x; D]$$

with C_p the principal value integral. Comparing (5.2) and (5.8), clearly integral operator $C[z; D]$ solves the normal form, so a particular solution is

$$(5.11) \quad C(z) = C[z; D] \equiv C[D]$$

with $C[D]$ an abbreviated notation for the Cauchy-type vector or matrix operator on density $D(x)$.

An illustration of these solutions is provided by the logarithmic normal forms in (4.10). General solutions follow as

$$(5.12) \quad \begin{aligned} \ln f(z) &= i\pi(I+2j)C[I] + \ln h_f(z) \\ \ln g(z) &= i2\pi k C[I] + \ln h_g(z) \end{aligned}$$

where h_f and h_g satisfying homogeneous logarithmic forms. Taking the matrix exponential (since the matrices are diagonal) gives

$$(5.13) \quad \begin{aligned} f(z) &= h_f(z) e^{i\pi(I+2j)C[I]} \\ g(z) &= h_g(z) e^{i2\pi k C[I]} \end{aligned}$$

The integral, $i2\pi C[I] = \int_L ds [s-z]^{-1}$ has logarithmic singularities at the ends of L , so the above exponentials give zeros or poles at endpoints, depending on the integers in diagonal matrices j and k . The $i\pi C[I]$ term in the expression for $f(z)$ yields square root branch points at the ends of L .

Consequently, $g(z)$ is a rational function with poles confined to the endpoints of L , and $f(z)$ has poles and square root branch points there.

Returning to the homogeneous normal form, the second of (4.4) shows that (5.4) is satisfied by any analytic function real on L . Reduction to a homogeneous matrix equation follows by factoring $H(z)$ as

$$(5.14) \quad H(z) = h(z) J$$

where J is the unit vector and $h(z)$ is diagonal, and substituting into (5.4), giving

$$(5.15) \quad (h_+ - h_-) J = 0$$

Without loss of generality it can be assumed that no component of H or h vanishes identically since a (nonvanishing) homogeneous solution can always be added. Therefore $h(z)$ satisfies the homogeneous matrix normal form and, from the second solution in (5.13), is a rational function with poles at the endpoints of L only, as is the resulting vector homogeneous solution, $H(z)=h(z)J$.

6. Dual Schwarz-type Integral Representations

Expressions for $W(z)$ are determined from dual boundary equations (4.4) by solving the general normal forms in (4.8). Replacing the right hand sides by the Cauchy integral factorization and solving for $W(z)$ yields the conjugate Schwarz-type integral representations

$$(6.1) \quad \begin{aligned} W(z) &= f^{-1}(z) \{ 2 C[z; f_+ U_0] + A(z) \} \\ W(z) &= g^{-1}(z) \{ i 2 C[z; g V_0] + B(z) \} \end{aligned}$$

where $A(z)$ and $B(z)$ are additive homogeneous vector solutions, and $f(z)$ and $g(z)$ are multiplicative homogeneous matrix solutions. At this point the order of endpoint behavior and other zeros in f and g is arbitrary, provided of course the integrals in (6.1) exist. In particular, the rational functions in f and g must be chosen to insure existence of these integrals and satisfy any explicit endpoint conditions given by $E(x)$, described in Sec. 3. Homogeneous solutions $A(z)$ and $B(z)$, in conjunction with $f(z)$ and $g(z)$, depend on the singularity function $S(z)$ and prescribed endpoint and infinity conditions. Methods for their evaluation are described in Sec. 8.

To insure consistency between the Schwarz-type representations it is necessary to examine boundary values of $W(z)$ in (6.1) from limits in (5.10), giving

$$(6.2) \quad \begin{aligned} W_{\pm} &= U_0 \pm 2 f_+^{-1} C_p[f_+ U_0] \pm f_+^{-1} A \\ W_{\pm} &= \pm i V_0 + i 2 g_+^{-1} C_p[g V_0] + g_+^{-1} B \end{aligned}$$

where $A_{\pm} = A$, $B_{\pm} = B$, and $g_{\pm} = g$ are real while $f_{\pm} = \pm f_+$ are imaginary on L . Since $W_{\pm} = U_0 \pm i V_0$, it follows that U_0 and V_0 are related as

$$(6.3) \quad \begin{aligned} i V_0 &= 2 f_+^{-1} C_p[f_+ U_0] + f_+^{-1} A \\ U_0 &= i 2 g^{-1} C_p[g V_0] + g^{-1} B \end{aligned}$$

which is a generalization of the Hilbert transform pair. These equations provide necessary conditions on boundary values of the conjugate harmonic vector functions. In applications, either the first or second of (6.1) is used to represent $W(z)$ in terms of U_0 or V_0 , and the respective conjugate boundary value is given by the first or second of (6.3). Note that eliminating either U_0 or V_0 between the Hilbert-type transform pair yields a double integral equation for each—generalizations of the Poincaré-Bertrand formula for repeated singular integrals.

7. Boundary Integral Equations

The boundary condition, (3.6), must be solved for U_0 or V_0 in order to evaluate $W(z)$ from one of the Schwarz-type boundary integral representations in (6.1). To this end, either the original form of the boundary condition can be considered, namely

$$(7.1) \quad a U_0 - b V_0 = \operatorname{Re}\{(a + ib)W_0\} = G$$

or alternatively, a vectorization of Riemann's classic form, obtained by replacing U_0 and V_0 from (4.4) giving

$$(7.2) \quad c_+ W_+ + c_- W_- = 2G, \quad c_{\pm} \equiv a \pm ib$$

Either expression can be used to reduce the boundary problem to a tractable boundary integral equation. Note that for convenience the \pm subscript on c denotes the function (+) and its complex conjugate (-), rather than boundary values of some analytic function (i.e., $c(z)$).

The obvious approach is to decouple boundary condition (7.1) by eliminating V_0 or U_0 using dual integrals (6.3), i.e., the Hilbert-type transform pair, resulting in a singular integral equation for each as

$$(7.3) \quad \begin{aligned} a U_0 + i 2 b f_+^{-1} C_p[f_+ U_0] &= G - i b f_+^{-1} A \\ i 2 a g^{-1} C_p[g V_0] - b V_0 &= G - a g^{-1} B \end{aligned}$$

Alternate forms of the integral equations can be written when a and b are invertible, by solving (3.6) for U_0 or V_0 and inserting directly in the Hilbert transform integrals. However, there is no obvious advantage except when a and b exhibit further degeneracy, e.g., they commute with inverses of g and f , respectively. A limiting example is Carleman's "algebraic" solution for scalar equations. Simplifications also arise when a and b are disjoint, i.e., share no common matrix elements.

A more constructive approach to reducing the boundary condition is to first extract an expression for $W_{\pm}(x)$ from Riemann's form (7.2). Factoring $2G$ on the right as

$$(7.4) \quad 2G = K_+ + K_-$$

yields the homogeneous form

$$(7.5) \quad (c_+ W_+ - K_+) + (c_- W_- - K_-) = 0$$

Therefore, the expression in parentheses is imaginary, i.e., $c_{\pm} W_{\pm} - K_{\pm} = \pm iR$ where $R(x)$ is a real function on L . Solving for W_{\pm} gives

$$(7.6) \quad W_{\pm} \equiv U_0 \pm iV_0 = c_{\pm}^{-1} (\pm iR + K_{\pm})$$

i.e., an explicit factorization for boundary values of $W(z)$ on L in terms of unknown real function $R(x)$. This reduction assumes that the inverse of c_{\pm} exists. Note that (7.4) is normalized and solved just like the first of (4.4), yielding a particular solution for K_{\pm} as

$$(7.7) \quad K_{\pm} = 2r_{\pm}^{-1} C_{\pm}[x; r_+ G] = G \pm 2r_+ C_p[x; r_+ G]$$

where $r(z)$ has square root branch points at the ends of L .

Boundary representation (7.6) is useful in its own right since it exhibits the singular behavior of U_0 and V_0 in terms of the given boundary data, namely, c_{\pm} . More to the point, either part, real or imaginary, can be continued off the boundary via Schwarz representation (6.1), provided of course that the corresponding Hilbert transform is satisfied. Therefore, this approach yields another conjugate pair of boundary integral equations by substituting the real and imaginary parts of (7.6) directly into the dual Hilbert equations. To examine the resulting integral equations, particular expressions for U_0 and V_0 follow from (7.6) as

$$(7.8) \quad \begin{aligned} U_0 &= -\text{Im}(c_+^{-1})R + \text{Re}(c_+^{-1}K_+) \equiv -qR + P \\ V_0 &= \text{Re}(c_+^{-1})R + \text{Im}(c_+^{-1}K_+) \equiv pR + Q \end{aligned}$$

which, when substituted into (6.3), give

$$(7.9) \quad \begin{aligned} if_+qR - 2C_p[f_+pR] &= 2C_p[f_+P] - if_+Q + A \\ gpR - i2C_p[gqR] &= i2C_p[gQ] - gP + B \end{aligned}$$

The principal advantage of these last integral equations over those in (7.3) is that unknown vector R is a residual factor, i.e., the principal boundary behavior is explicitly factored out, whereas it is implicit in U_0 and V_0 . Therefore, a "smoother" basis can be used in solving (7.9) for factor R . This assumes of course that the inverse of $c_+ = a + ib$ exists, and if not, (7.3) provides the necessary equation. Therefore, to solve the boundary problem in terms of either Schwarz-type representation in (6.1), it is necessary to solve one of the singular integral equations in (7.3) or (7.9). Of course, the resulting solutions are only unique to the extent that homogeneous solutions A , B , f , and g are known.

8. Determination of Homogeneous Solutions

The general Schwarz-type representations obtained for $W(z)$ in Sec. 6 are determined there to within certain homogeneous solutions, namely vectors $A(z)$ and $B(z)$, and diagonal matrices $f(z)$ and $g(z)$. These must be evaluated on the basis of all remaining information on the problem, namely, the prescribed singularity function, $S(z)$, endpoint behavior and convergence requirements, and properties of the line of integration, L .

$S(z)$ prescribes the isolated singularities of $W(z)$ off L in the upper half-plane and is regular on L by hypothesis. Evaluating the residue of $W(z)$ from Schwarz representation (6.1) at an arbitrary isolated singularity, say z_0 , yields

$$(8.1) \quad r_0 \equiv \text{Res}\{W(z)\} = \text{Res}\{S(z)\} = \text{Res}\{f^{-1}(z) A(z)\} = \text{Res}\{g^{-1}(z) B(z)\} \quad \text{at } z_0$$

assuming that $f(z)$ and $g(z)$ do not vanish at z_0 . Thus, in principle $A(z)$ and $B(z)$ can be obtained from $S(z)$. For example, without loss of generality $S(z)$ can be defined as a sum composition of singularities with complex conjugate images in the lower half-plane, thereby satisfying the reflection principal, $S^*(z)=S(z^*)$. In the case of a simple pole $A(z)$ and $B(z)$ follow directly from this continuation after multiplying the upper half-plane residues by matrix constants $f(z_0)$ and $g(z_0)$ respectively, and residues in the lower half-plane by $f^*(z_0)$ and $g^*(z_0)$. For higher order poles A and B can still be determined from S , in principle, although the residue calculus becomes more involved.

To evaluate rational functions in $f(z)$ and $g(z)$ given by (6.2), it is necessary to examine the homogeneous solution of Riemann's form for $W_{\pm}(x)$, given by (7.11) with $G \equiv 0$,

$$(8.2) \quad W_{\pm} = U_0 \pm i V_0 = \pm i (a \pm i b)^{-1} R$$

Although real and imaginary parts do not satisfy the Hilbert transform pair in general, this expression does exhibit the singular behavior of $W_{\pm}(x)$ explicitly in terms of boundary coefficient $c_{\pm}^{-1} = (a \pm i b)^{-1}$. In particular, when the determinant of c_{\pm} vanishes on L , then U_0 and V_0 have corresponding infinities for which f and g must have zeros of the proper order to insure convergence of the integrals in (6.1). Therefore, homogeneous factors $f(z)$ and $g(z)$ provide means for factoring out irregularities in the density function of the Cauchy-type integrals, both at interior points and endpoints on the line of integration.

Additional conditions are provided by the point at infinity and the prescribed endpoint order function, $E(z)$. Depending in part on the mapping of the original domain to the half-planes, the behavior at infinity is regular or singular and the homogeneous terms must exhibit it. This behavior is known a priori and typically incorporated by an order condition in the endpoint function as $x \rightarrow \pm\infty$. Consequently, in addition to removing singularities in the density functions, multiplicative homogeneous solutions $f(z)$ and $g(z)$ are chosen to incorporate these prescribed endpoint conditions. Of course, all of the conditions must be self-consistent and admit solutions of

the boundary condition singular integral equations derived in Sec. 7.

9. Degenerate Solutions of the Riemann-Hilbert Form

An alternative to solving integral equations is to examine the more familiar (historically) Riemann-Hilbert form of boundary problem. The starting point is Riemann's form in (7.2), which yields a generalization of Hilbert's form as

$$(8.1) \quad W_+ = -c_+^{-1} c_- W_- + 2 c_+^{-1} G$$

assuming c_{\pm}^{-1} exists. This vector form cannot be normalized in general since c_{\pm} has off-diagonal terms that, after continuation off L , mix the complex variable dependency. This violates the implicit vector convention that the n^{th} component of W is a function of z_n only, i.e., $W_n = W_n(z_n)$. However in a degenerate case—isomorphic to the scalar theory—the Riemann-Hilbert form can be normalized and solved.

An examination of Hilbert's form (8.1) shows that when $c_+^{-1}c_-$ is diagonal the problem reduces to a logarithmic normal form. In this case introducing factors d_{\pm} such that

$$(8.2) \quad d_+^{-1} d_- = d_- d_+^{-1} = -c_+^{-1} c_-$$

where $d(z)$ is a diagonal matrix function, and taking the principal value logarithm gives

$$(8.3) \quad \text{Ln } d_+ - \text{Ln } d_- = \text{Ln } (-c_+^{-1} c_-)$$

with particular solution

$$(8.4) \quad d(z) = s(z) e^{C[z; \text{Ln}(-c_+^{-1} c_-)]}$$

where $s(z)$ satisfies the homogeneous normal form. Replacing $-c_+^{-1}c_-$ in Hilbert's form using (8.2) and multiplying by d_+ gives the normal form

$$(8.5) \quad d_+ W_+ - d_- W_- = 2 d_+ c_+^{-1} G$$

with general solution

$$(8.6) \quad W(z) = d^{-1}(z) \{2 C[z; d_+ c_+^{-1} G] + M(z)\}$$

where homogeneous solution $M(z)$ follows from the singularity function. Of course, (8.6) is a degenerate solution since it is only valid when $c_+^{-1}c_-$ is a diagonal matrix.

10. Discussion

The foregoing analysis has developed a vectorized theory for multiply connected boundary problems involving vector unknowns defined either in a single domain or in connected domains. The terminology, multiply connected, means that two or more functions in one or more domains are coupled at the limit points. In general, the problems are mapped to several complex half-planes

(canonical domains), each supporting one component of the unknown, with a linear matrix equation (boundary conditions) relating real and imaginary parts of all coupled components on the real axes. The half-planes are analytically continued by reflection, giving domains with branch cuts on the real axes. These mapped and continued domains are similar to Riemann sheets, but here each is defined by a unique complex variable in a diagonal matrix, rather than by one branch of a multiple-valued function. The somewhat pretentious term, parallel complex planes, is introduced to distinguish this case—indicating in particular that domains meet at their limit points only, either continuous (the real axis) or isolated. The only functional connection is across these limit points, which consist of mapped boundaries, branch cuts, and isolated singularities.

The classical approach to scalar problems (single complex variable) is the Riemann-Hilbert formalism, due in part to Carleman. However, the present vector form requires a more fundamental theory based on continuation of the Schwarz boundary equations off the real axes. This is developed using so-called Schwarz normal forms and their algebraic or logarithmic solutions given by vectors of Cauchy-type integrals over appropriate segments of the axes. The integrals follow from the sifting property of the delta function and its continuation off the real axes. These solutions provide a general vector basis for representing analytic functions in several complex planes using integrals of real or imaginary parts. The representations explicitly yield the Hilbert-type transform expressing consistency (conjugacy) between real and imaginary boundary values.

With the analytic Schwarz-type representation available, the boundary problem can be reduced to a tractable form. The most direct approach is to decouple the boundary condition by eliminating either the real or imaginary vector using the Hilbert-type transform pair—resulting in two vector singular integral equations, only one of which need be solved for either the real or imaginary part. A more constructive approach is to extract a boundary factorization (with analytic unknown) from Riemann's form of the boundary condition and substitute the real or imaginary part into the Hilbert-type transform pair, yielding two singular integral equations in terms of a "smoother" unknown.

Due to off-diagonal terms in the coefficients, closed-form solutions of the singular integral equations from either approach do not appear possible in this vector case. However, discrete (approximate) algebraic solutions of the integral equations may be found by replacing the Cauchy-type integrals by open-ended quadrature formulas and solving. Such vector quadrature formulas present no new difficulties, the only drawback being size of the resulting algebraic system and its inversion. When boundary condition matrix coefficients a and b or $c \equiv a + ib$ are invertible, a recursive solution of the integral equations is possible, thereby eliminating the need for a matrix inversion.

A general solution of the boundary problem is found by evaluating the Schwarz integral of the real or imaginary part. This evaluation also reduces to a quadrature formula in practice. To

complete the general solution, as well as uniquely solving either integral equation derived from the boundary condition, certain homogeneous solutions must be evaluated. They arise since the Schwarz-type representations are only unique to within multiplicative and additive analytic functions. These functions are determined by requiring that all Cauchy-type integrals be convergent, and by including any endpoint behavior or isolated singularities prescribed a priori. Convergence of the integrals depends on removing nonintegrable singularities derived from the complex boundary coefficient matrix, $c = a + ib$.

Some attention has been given to the Riemann-Hilbert approach to these problems in order to show its limitations and justify the procedure developed above. The classical Riemann-Hilbert formalism in one complex variable typically avoids the solution of singular integral equations by applying Carleman's algebraic reduction and writing inversion formulas directly. It is not generally applicable to the vector problem since there are off-diagonal terms in the matrix coefficients. However, in a degenerate case the formalism does apply, namely, when the boundary coefficients are diagonal. This vector case does not appear useful in general, but instead completes the connection between scalar and vector theories.

11. Conclusions

The principal result of this paper is a constructive solution procedure for vector boundary problems in two-dimensions, i.e., problems involving two or more unknown harmonic functions coupled on the limit points of one or more plane domains. The vectorized theory offers a significant analytical advantage over iterative methods such as Schwarz's alternating procedure and similar schemes. The mathematics reads like a scalar theory but is uniformly valid for any number of harmonic vector components.

The vectorized theory generalizes the Riemann-Hilbert formalism for scalar boundary problems at the expense of Carleman's algebraic technique—because the equivalence between Riemann-Hilbert problems and singular integral equations in one complex domain is a degenerate case. It is not valid for the vector problem since Carleman's approach, based essentially on a product factorization of the coefficient in Hilbert's form, fails for matrix coefficients with off-diagonal terms—even if commuting factors exist.

The analysis has considered so-called canonical domains, i.e., half-planes. These are straightforward analytically and appropriate for discontinuous boundaries (finite or semi-infinite lines of integration). However, for continuous boundaries (infinite lines) it is often better to map the original domains directly to interiors of circles and trade the functional simplicity of Cauchy kernels and line integrals for their circular equivalents, namely, Hilbert kernels and contour integrals. It is a simple exercise to generalize all of the paper's results to the case of circular contours. Boundary continuity also bears on the apparent freedom to choose either of two

Schwarz-type representations for $W(z)$, either on the real or imaginary boundary value. The real kernel exhibits a multiplicative function (homogeneous solution) with square root branch points at endpoints and is appropriate for finite or semi-infinite lines of integration; while the imaginary kernel does not admit multiplicative branch points and is best suited to infinite lines, corresponding to continuous boundaries.

This analysis has not examined with any rigor the question of existence and uniqueness of solutions. Rather, emphasis has been on vectorization and generalization of well-known methods to construct an analytical basis for reducing vector boundary conditions to a tractable form. The crux of the analysis is the pair of vector singular integral equations derived from the boundary conditions, and their solution. Existence and uniqueness depend on order or continuity conditions at endpoints and the point at infinity (governed by the boundary condition matrices and mappings of the original domains) and the multiplicative and additive homogeneous solutions (determined by enforcing convergence of the Cauchy-type integral representations and incorporating prescribed poles). These aspects require further study in general, particularly in the context of real boundary data.

As mentioned in the Introduction and exemplified in Sec. 2, this theory provides a constructive basis for solving some interesting two-dimensional vector diffraction problems that are unsolvable by classical analysis methods. The problems involve multiple wedge-shaped domains connected on their faces from a common vertex. Each wedge supports one or more wave functions coupled by nonseparable continuity conditions across the faces. By considering homogeneous solutions of degree zero, i.e., self-similarity, and a so-called characteristic transformation, these problems yield the vector boundary problem studied here. The diffraction problem is thereby reduced to characterizing the boundary condition matrices obtained for the various wave functions and boundary conditions considered, and solving the resulting singular integral equations. Electromagnetic and elastic waves are of particular interest, with applications ranging from optical diffraction in the small (integrated optics) to seismic diffraction in the large (crustal seismology).

In addition to the diffraction problem there are certainly a variety of multiply connected harmonic boundary problems for which the technique is applicable, although many of these have no doubt been formulated and solved using other approaches. Of more practical interest is the future extension of this theory to multiply connected biharmonic problems, particularly in the two-dimensional theory of elasticity, e.g., [4] and especially [5]. Examples of interest include heating of a finite multi-layer strip (e.g., bi-metal), and elastic contact problems involving a deformable (rather than rigid) indenter.

REFERENCES

- [1] T. Carleman (1922). *Sur la résolution de certaines équations intégrales.*, Arkiv för Matematik Astronomi och Fysik, Bd. 16, No. 26.
- [2] R. Courant and D. Hilbert (1962). *Methods of Mathematical Physics*, Vol. II, Interscience Publishers.
- [3] D. Hilbert (1904). *Über eine Anwendung der Integralgleichungen auf ein Problem der Funktionentheorie.* Verhandl. des III. International Mathematiker Kongresses, Heidelberg.
- [4] S. G. Mikhlin (1957). *Integral Equations*, Pergamon Press, New York.
- [5] N. I. Muskhelishvili (1935). *Some Fundamental Problems of the Theory of Elasticity*, Noordhoff International Publishing.
- [6] N. I. Muskhelishvili (1944). *Singular Integral Equations*, Noordhoff International Publishing.
- [7] B. Riemann (1953). *Gesammelte Mathematische Werke*, Dover Publications, Inc., New York.
- [8] Wojcik, G. L. (1989). 'A Numerical and Theoretical Study of Seismic Wave Diffraction in Complex Geologic Structure,' Final Report AFGL-TR-88-0291, Air Force Geophysics Laboratory Contract F19628-84-0102, Weidlinger Associates.