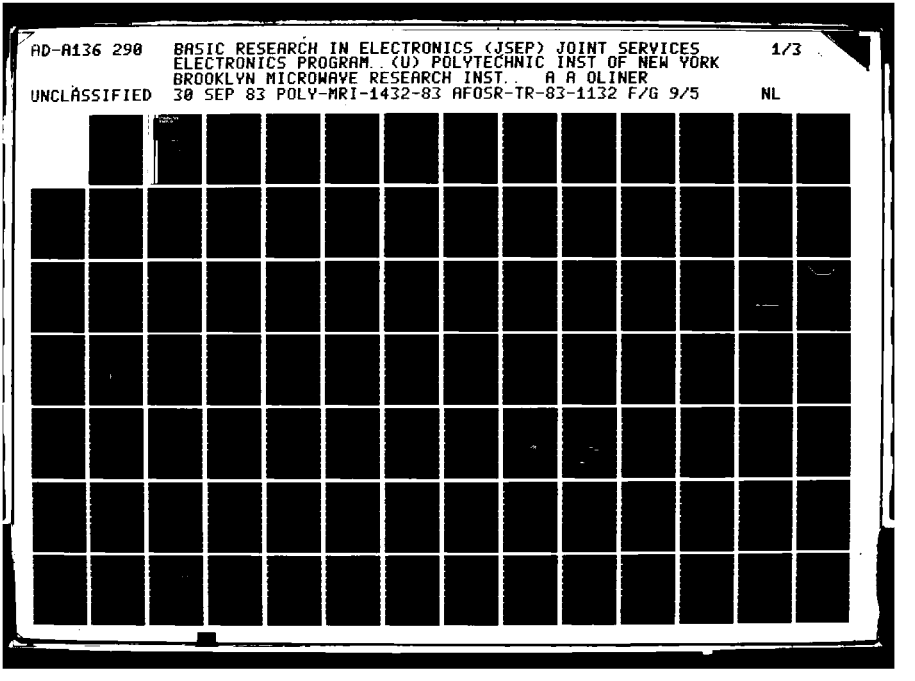
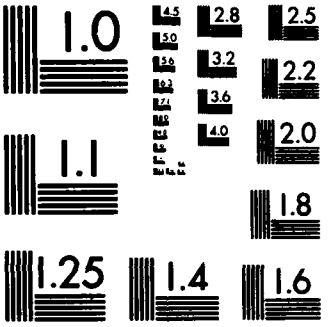
AD-A136 290 BASIC RESEARCH IN ELECTRONICS (JSEP) JOINT SERVICES ELECTRONICS PROGRAM..(U) POLYTECHNIC INST OF NEW YORK BROOKLYN MICROWAVE RESEARCH INST.. A A OLINER 1/3

UNCLASSIFIED 30 SEP 83 POLY-MRI-1432-83 AFOSR-TR-83-1132 F/G 9/5 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Polytechnic Institute of New York

AFOSR-TR- 83-1132

30 SEPTEMBER 1983

ANNUAL REPORT

on

BASIC RESEARCH IN ELECTRONICS (JSEP)

CONTRACT F49620-82-C-0084

APRIL 1, 1982 TO MARCH 31, 1983

POLYTECHNIC INSTITUTE OF NEW YORK
MICROWAVE RESEARCH INSTITUTE
BROOKLYN, NEW YORK 11201

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR-TR-83-1132 / 2. GOVT ACCESSION NO.: AD-A136290	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Basic Research in Electronics (JSEP)	5. TYPE OF REPORT & PERIOD COVERED: ANNUAL REPORT 1 April 1982-31 March 1983
	6. PERFORMING ORG. REPORT NUMBER: POLY-MRI-1432-83 ✓
7. AUTHOR(s): Arthur A. Oliner	8. CONTRACT OR GRANT NUMBER(s): F49620-82-C-0084
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Microwave Research Institute, Polytechnic Institute of New York, Brooklyn, New York 11201	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NE, Bolling Air Force Base, Washington, D.C. 20332	12. REPORT DATE: 30 September 1983
	13. NUMBER OF PAGES: 256
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Electromagnetics, microwaves, millimeter waves, waveguides and antennas, optics, x-rays, solid state interactions, solid state materials, surface acoustic waves, information electronics, systems, image restoration

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Annual Report presents a summary of the scientific progress and accomplishments on research projects funded by the Joint Services Electronics Program (JSEP) for the contract period from 1 April 1982 through 31 March 1983. It does not contain information regarding accomplishments on research projects funded in other ways.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

The Joint Services Electronics Program at the Polytechnic is the core of interdisciplinary research in electronics encompassing programs in the Departments of Electrical Engineering and Physics under the aegis of the Microwave Research Institute. The research encompassed by this program is grouped under three broad categories: Electromagnetics, Solid State Electronics and Information Electronics. The detailed projects (work units) comprising the complete program are listed in the Table of Contents.

In addition to the progress, we include for each of the work units the publications during the past year resulting from the research, and some of the recent interactions with the DoD or industry or other academic institutions. At the back of this Final Report we present our Report on Significant Scientific Accomplishments.

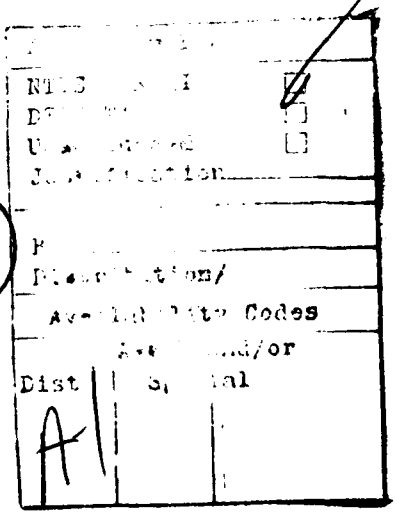


TABLE OF CONTENTS

Polytechnic Institute of New York
Microwave Research Institute

ABSTRACT

This Annual Report presents a summary of the scientific progress and accomplishments on research projects funded by the Joint Services Electronics Program (JSEP) for the contract period from 1 April 1982 through 31 March 1983. It does not contain information regarding accomplishments on research projects funded in other ways.

The Joint Services Electronics Program at the Polytechnic is the core of interdisciplinary research in electronics encompassing programs in the Departments of Electrical Engineering and Physics under the aegis of the Microwave Research Institute. The research encompassed by this program is grouped under three broad categories: Electromagnetics, Solid State Electronics and Information Electronics. The detailed projects (work units) comprising the complete program are listed in the Table of Contents.

In addition to the progress, we include for each of the work units the publications during the past year resulting from the research, and some of the recent interactions with the DoD or industry or other academic institutions. At the back of this Annual Report we present our Report on Significant Scientific Accomplishments.

SECTION I
ELECTROMAGNETICS

SECTION I: ELECTROMAGNETICS

A. NEW PHYSICAL EFFECTS INVOLVING OPEN DIELECTRIC STRUCTURES

Professors A.A. Oliner and S.T. Peng

Unit EM3-1

1. OBJECTIVE(S)

This study is concerned with guiding, radiating and scattering effects involving open dielectric structures. Almost all of the published literature on these topics involves a surface wave incident on a dielectric junction or discontinuity or grating, etc., at normal incidence, where the resulting boundary-value problem is two-dimensional, and TE and TM modes do not couple to each other. In connection with earlier investigations of open dielectric waveguides and antennas for millimeter-wave integrated circuits and for integrated optics, it was necessary for us to examine the behavior of dielectric structures for which the surface wave is obliquely incident. The boundary-value problems then change from the scalar two-dimensional ones to vector three-dimensional ones, and the TE and TM modes no longer remain independent but are coupled together. The vector nature of these problems introduces new mathematical challenges, but, more interestingly, we found that the mode coupling produced a rich variety of interesting and sometimes unexpected new physical effects.

In this contract, we continue to explore these physical effects on old and new types of dielectric structure. It is, of course, necessary to first obtain the appropriate mathematical solutions; then we examine the physical consequences of those solutions, and, where appropriate, assess their implications for device performance in millimeter-wave integrated circuits and integrated optics.

2. APPROACH

As mentioned in the section above, the analytical challenges involve vector three-dimensional problems which include TE-TM mode coupling. In addition, the presence of the TM modes produces very slow convergence, thus necessitating additional mathematical tricks to help accelerate the convergence. We have already produced a (novel) rigorous solution for a surface wave guided obliquely along a dielectric grating, by appropriately combining the solutions for two constituent scalar problems, and we have introduced a matrix partition procedure to speed up the mode convergence in the scattering by a dielectric step junction. All of the approaches we employ, both novel and standard, fall into the general category of guided wave techniques; we intend to continue to explore new techniques within this framework to obtain better solutions for existing problems and new solutions for new problems.

Since the basic goal in this study is the identification and understanding of new physical effects, it is best if we have available a rigorous solution so that effects are not missed, and spurious results do not arise, due to approximations which are not fully understood. On the other hand, in most cases these rigorous results are quite complicated, requiring an involved computer program, particularly when complex roots are involved, as they often are. Therefore, after the

physical effects are understood, and their implications for device performance assessed, we propose to derive simple approximate formulations, preferably in closed form, so that the magnitudes of these physical effects can be calculated quickly and easily. In addition, we expect to perform experiments at microwave or millimeter-wave frequencies to verify some of the physical effects.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contribution can be understood more clearly.

1) We were the first to show theoretically that certain dielectric strip waveguides for integrated optics or for millimeter wave integrated circuits can leak under appropriate circumstances, producing undesired cross talk between neighboring components. The only published experimental confirmation of that effect was indirect in nature. We have this past year made measurements which constitute a direct experimental confirmation of these leakage effects, and we have presented a paper[40] at a symposium on this work. It is also interesting that some of the measurements were made in Japan by a collaboration we have established with a professor there; the paper was presented jointly with him. A joint journal article is in preparation.

2) In last year's progress section it was indicated that a resonance process occurs in connection with the leakage effects on dielectric strip waveguides mentioned above. Also, a "debate" has appeared in the literature on whether this resonance effect results in nulls or simply sharp and deep dips. Last year, we devised a proof that the sharp dips were not actually nulls, and indicated that the proof would be included as a part of a doctoral thesis to be completed. An improved form of the proof has been devised during this past year, and it has been included in the thesis[32] which was completed in June 1983. The writing of a journal publication containing this proof, together with the background, will be started soon.

3) We had previously obtained careful experimental data from Professor R. Ulrich of the University of Hamburg-Harburg, West Germany, on dispersion values in the neighborhood of the Bragg reflection region for a periodic optical waveguide. That structure is composed of a periodic grating of dielectric grooves on a dielectric layer on a dielectric substrate, where the surface wave is incident at an angle on to the grooves, and the structure is of potential importance for multiplexers and other devices for integrated optics. When the surface wave is incident at an angle, as here, four stop bands are obtained in the dispersion plot instead of the usual two. We developed a rigorous theory two years ago for such situations, and we previously compared our theory with Ulrich's measurements, and found excellent agreement in general shape and even reasonable agreement in numbers. Curves showing both the measurements and theory were presented previously as Fig. 4, but we believed that the discrepancy arose because our theory applied to rectangular grooves, whereas their measurements were made on grooves which were more sinusoidal than rectangular. Last

September, on a side trip associated with attendance at the European Microwave Conference, Professor Oliner visited Professor Ulrich in Hamburg, and spent two days with him discussing these earlier measurements and planning future collaboration (see item (4) below). It turned out that the dimension given to us earlier by letter for the layer thickness was the value before the grooves were etched, and that an SEM photo of the structure showed that the grooves were primarily rectangular, not sinusoidal. When we recalculated our theoretical values using the new value for layer thickness, remarkable agreement with the earlier measurements was obtained. A new Fig. 4, which replaces the old one, now shows this gratifying agreement.

4) Part of the planning for future collaboration with Professor Ulrich, referred to in (3) above, involved a series of suggestions for us to perform certain calculations of the Bragg stop band widths as a function of particular geometric parameters, which Ulrich believed would reveal some additional physics in the stop band behavior, as well as be of value to designers of optical multiplexers using the Bragg reflection technique. During this past year, we made such calculations, and they were presented (together with the new agreement found with Ulrich's earlier measurements mentioned in (3) above) jointly with Ulrich at a symposium[41] in June 1983. A journal paper on this work is planned.

5) One of the important new physical effects introduced by oblique, rather than normal, surface wave incidence on a periodically-grooved dielectric layer is that of cross polarization, due to the TE-TM wave interactions occurring at every groove step when the incidence angle is oblique. Studies of cross polarization were begun a bit over a year ago, but during this past year we have examined carefully the nature and extent of such cross polarization when the periodic structure is designed to radiate. We expected a significant amount of cross polarization to be present, but we were actually surprised by the strength of it. For normal incidence on the grooves there is no cross polarization, of course, but the amount becomes quite strong even for small deviations from normal. For an incident TE surface wave, for example, the TM content of the radiated power actually exceeds the TE content for angles from the normal greater than 15° or so. For an incident TM surface wave, even stronger cross polarization effects are found; for example, in one case, for an angle somewhat less than 15°, a maximum occurs at which about 90% of the radiated power consists of the "wrong" polarization. Some of these results were presented at a symposium[42] in May 1983. There are also implications here for millimeter wave antennas of finite width, where the constituent surface waves cross the grooves at an angle; these aspects are being pursued.

6) The derivation of the theoretical results for the guided wave behavior of surface waves on these periodically grooved dielectric layers involves a transverse resonance procedure, which is the same as saying a self-consistency condition in the transverse plane. In this problem, the self-consistency involves the multimode scattering of plane waves incident at an angle on the grooves. Thus, a constituent part of the guided wave problem is the subsidiary problem of the scattering produced by a plane wave of arbitrary polarization content incident at an arbitrary angle in both orthogonal directions on a periodically grooved dielectric layer. That subsidiary problem was investigated independently during this past year, and some new results were found, including

resonance effects and polarization interchanges. These results were presented at a symposium[43] in August 1983, and some of the new features were included in a recently completed Ph.D. thesis.[37]

7) When a number of these new physical effects are placed together, they form a cohesive and very interesting package. With this viewpoint in mind, we prepared an overview paper which we presented at a symposium[44] in August 1983. The presentation was very well received, with many compliments.

8) The investigations so far have concentrated on effects involving dielectric strip waveguides and periodically grooved dielectric layers with obliquely incident waves. Although such investigations are not complete, we have also begun an examination of a somewhat more complicated structure: a dielectric layer grooved periodically in two orthogonal directions, resembling a waffle iron. The reason for the new investigation is that our earlier brief look into the behavior of such a structure indicated that an enormous variety of weird-looking wavenumber diagrams may be possible. The implication is that additional new physical effects may emerge from such a study. That look was performed "kinematically," i.e., by simply examining various possibilities graphically, employing the various unperturbed interacting wavenumber curves. Recently, we have begun an analytical investigation, taking a model in which the variations in one direction are independent of those in the other orthogonal direction. We have only just begun, but we already see that some of the crossings that we assumed earlier to be interacting are in fact not; certain additional wavenumber conditions must be satisfied before the interactions are physically real ones. There remains, however, a large number of interesting-looking diagrams that clearly warrant a further look into the behavior of such structures.

4. STATE OF THE ART AND PROGRESS DETAILS

As the earlier studies progressed, it became clear that the methods of analysis and, in many cases, the basic physical behavior of dielectric waveguides for integrated optics and for millimeter-wave integrated circuits are quite similar and have much in common, although important minor differences persist. In addition, it became evident that before the operation of these waveguides and antennas could be understood completely it was necessary to solve certain basic canonical guiding and scattering problems involving open dielectric structures. The solutions in the literature for these canonical problems were either non-existent or only approximate. Our attempts to solve these problems revealed that the problems were vector in nature, whereas almost everything previously available was only scalar, and that the TE and TM modes necessarily became coupled together. In the course of the analyses, we were fortunate enough to recognize, where others did not, that the TE-TM coupling gives rise to a variety of interesting physical effects, some of them unexpected, and that these effects have implications for device performance. The present contract builds on these accomplishments and continues them.

Before presenting various details in connection with these canonical problems and the new physical effects associated with them, we summarize briefly the principal accomplishments so far in that context. Four basic dielectric structures are considered: a dielectric step junc-

tion, a dielectric strip waveguide, a planar periodically-grooved dielectric layer, and a dielectric grating of finite width placed on a dielectric image guide (of the same finite width).

(i) Dielectric step junctions

We were the first to present a solution for the scattering produced by an obliquely-incident surface wave. That solution was approximate but it contained all the physics. It showed, for instance, that an incident wave gives rise to reflected and transmitted waves of the opposite polarization; in addition, for TE surface waves the junction exhibits a Brewster angle, which we were first to point out.[1] When we improved the solution numerically via a rigorous mode-matching procedure, we discovered the existence of basic convergence difficulties, and we developed a matrix partition scheme[2] to accelerate the convergence.

(ii) Dielectric strip waveguides

We were the first to show that under appropriate conditions most modes on a large class of these waveguides are leaky,[3] instead of being purely bound, as was expected by everyone. We showed that resonance effects[3] appeared in the leakage process. Furthermore, these leaky modes constitute a new class of leaky modes, since the leaking energy possesses a polarization opposite to that present in the main body of the waveguide, in contrast to customary leaky modes. Later, we obtained more accurate numerical results by a rigorous mode-matching procedure,[4-7] and verified that the basic physical effects determined earlier were correct. We developed a simple graphical procedure[8] to determine when leakage will occur, and simple relations for the angles of leakage. We applied the general analysis to various waveguides for integrated optics and for millimeter-wave integrated circuits, and found that for some waveguides the leakage rate is very small and that for others it can be enormous. We also described the implications for cross-talk between neighboring circuits and presented a novel directional coupler[9] based on the leakage effect. Recently, measurements[10] were made in Japan based on our theoretical predictions which verified the validity of our theory. In recognition of the fundamental accomplishments described here, we were invited to present a pair of papers[11,12] which keynoted a special issue on open waveguiding structures.

(iii) Oblique guidance by planar periodically-grooved dielectric layers

We are the only ones so far to have derived a rigorous theory,[13] and to have developed the corollary computer program, for oblique guidance by surface waves along a periodically-grooved dielectric surface. All previous solutions apply only for the scalar case of normal incidence. The results for oblique guidance are found to exhibit a large variety of physical effects which are not there at all when the guidance is normal to the grooves. One example involves the presence of four stop bands[14,15] instead of the usual two, in the Bragg reflection region; our theory also exhibits excellent agreement[15] with recent careful measurements by Ulrich[14] in that region. We have also observed that a number of interesting additional physical phenomena should be present, although we have not yet been able to investigate them: anisotropy phenomena, leading to beam steering which sometimes varies

rapidly with small frequency or angle changes, and beam focusing effects despite the straight line nature of the grating grooves; radiation of beams at quite peculiar angles; strong cross-polarization effects in portions of the radiation region; etc. Some of these effects were presented by us in a recent invited talk.[16] None of these effects has been noted previously, except for the beam steering.[14]

(iv) Dielectric gratings of finite width

At this stage, we have only observed that new physical effects are introduced by the finite width of dielectric gratings, and that these effects have important implications for device performance; these investigations have only just begun. Dielectric waveguides for millimeter waves possess narrow widths so as to permit single-mode operation; gratings are placed on them to produce either highly-reflecting structures for resonators or filters, or else to create leaky-wave antennas. We have found that the effect of finite width on the reflecting grating is to produce an extra (unwanted) transmission dip or reflection peak near to the desired peak, and therefore to seriously alter the performance.[15] No such problem occurs on similar SAW resonators where the grooves are very wide. On radiating gratings, an unexpected effect of finite width is to produce cross-polarized radiation. At this time, it is unclear whether the cross-polarized component is large or small over most of the scan range, but we expect that in a band near broadside such effects should be significant.[17] The important point to note now is that the effects mentioned above have not yet been looked for by experimenters and are quite unexpected. As a byproduct of this investigation, we have adapted the analytical result mentioned above for oblique guidance by a dielectric grating to obtain, in simple form, a theory for radiation from a leaky-wave antenna composed of a grating on a dielectric image guide (which is finite in width). This theory[18,19] yields the α and β of the leaky wave, and thus permits one to design such an antenna in accordance with desired radiation characteristics. It is the only such theory available to date.

As indicated above, these investigations were motivated by developments in dielectric waveguides for integrated optics and for millimeter-wave integrated circuits, and by grating antennas placed on waveguides for millimeter waves. The antennas had not been analyzed at all, and only early measurements of various types were available. For the waveguides, however, approximate theoretical treatments[20-25] appeared in the literature which were based on either the Marcatili procedure[26] (primarily for the optical waveguides) or the effective dielectric constant method[23,27] (mainly for the millimeter-wave structures). These approximate treatments assumed that only the dominant surface wave mode need be included in the analysis, and they also neglected entirely the step junctions at the sides of the waveguide. We sought to perform a better analysis and to assess the validity of these approximate solutions. Towards this goal, we simultaneously examined the strip waveguides directly, and also the dielectric step junction, corresponding to one side of a strip waveguide, as a canonical discontinuity problem relevant to waveguides and also of interest in its own right. First, we permitted the presence of both TE and TM surface waves in the analysis (instead of only one of them), and accounted approximately for the junction discontinuities; later, we took all higher modes into account as well, producing rigorously-phrased results. It

turned out that essentially all of the new physics emerged during the first of these stages (permitting the TE-TM mode coupling at the discontinuity); the inclusion of all the higher modes, and the rigorous mode matching at the junction, provided more accurate numerical values and verified the physics, but essentially did not add any new physics.

In order to help the reader to understand better the nature of our contributions and, in particular, the new physical effects which arise due to TE-TM mode coupling, we presented in the latest major proposal (EM2-1) the physical ideas underlying the scattering and guiding behavior of the various canonical structures mentioned earlier. In the present proposal we omit most of the details associated with those presentations, but we retain such material as is necessary to place into perspective the new accomplishments during the past year. The stress in the remaining discussion is therefore placed on progress during the past year, which occurred in connection with sections B, C and D below.

A. Surface Wave Incidence on a Dielectric Step Junction

During the past year we have not done anything further on this topic.

B. Guidance and Leakage Properties of Dielectric Strip Waveguides

Examples of dielectric strip waveguides for integrated optics and for millimeter-wave integrated circuits (almost all of which we have analyzed) are shown in Figure 1. Dielectric strip waveguides may be viewed as composed of two dielectric step junctions with a length of planar dielectric waveguide between them. Propagation along such waveguides may then be described in terms of a surface wave which bounces back and forth between the sides, or step junctions, undergoing "total reflection" at each bounce. Thus, the interesting effects associated with a single step junction should apply here.

Let us first recognize that the approximate theories, such as the Marcatili procedure[26] and the effective dielectric constant method,[23,27] assume that only one surface wave mode is present and that the geometrical discontinuities at the step junctions can be neglected. These approximate methods, which have been widely applied[20-24] to the structures in Fig. 1, thus neglect the TE-TM mode coupling occurring at the sides of the waveguides, and they thus miss the interesting physical effects. Such coupling produces a leaky mode instead of the purely bound modes predicted by the approximate theories.

The situation which gives rise to leaky modes is summarized in Figure 2. Without the TE-TM mode conversion at the sides of the waveguide, only the TE mode is considered to be present by the approximate theories. That assumption requires that the incident and reflected TE "rays" in Fig. 2 are above cut-off (propagating), whereas the transmitted "ray," in the outside region, is below cut-off (evanescent), and that no TM mode contributions are present. In the more accurate situation, where mode coupling is included, TM mode "rays" must be added, as seen in Fig. 2. Since we are depicting the situation for which the TM mode is dominant, the reflected TM wave inside is certainly propagating but the transmitted TM wave outside may be propagating or evanescent, depending on the geometrical parameters.

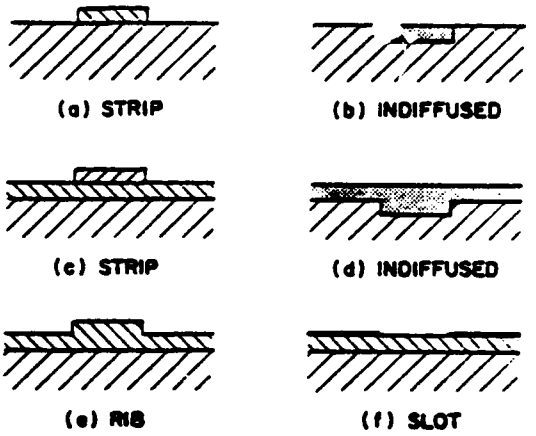


Fig. 1(a) Examples of open dielectric waveguides for integrated optics.

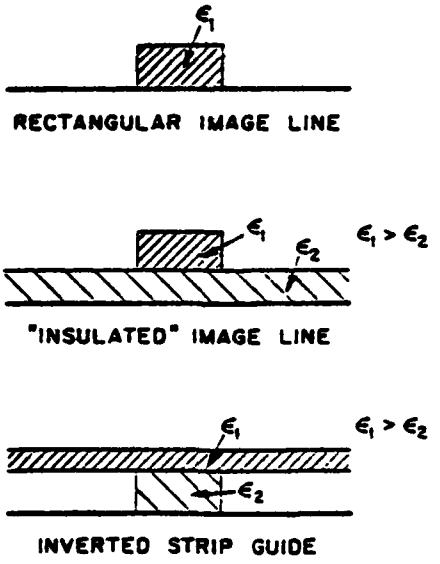


Fig. 1(b) Examples of open dielectric waveguides for millimeter waves.

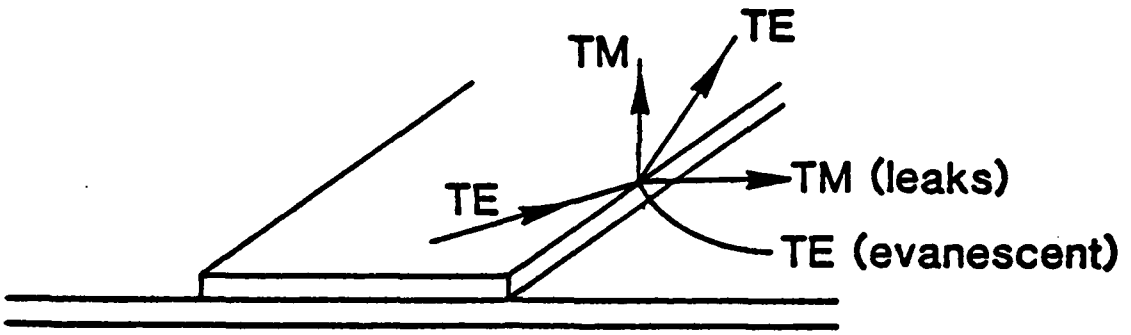


Fig. 2 Pictorial representation of mode-conversion effects at the side of a dielectric waveguide that give rise to leaky modes.

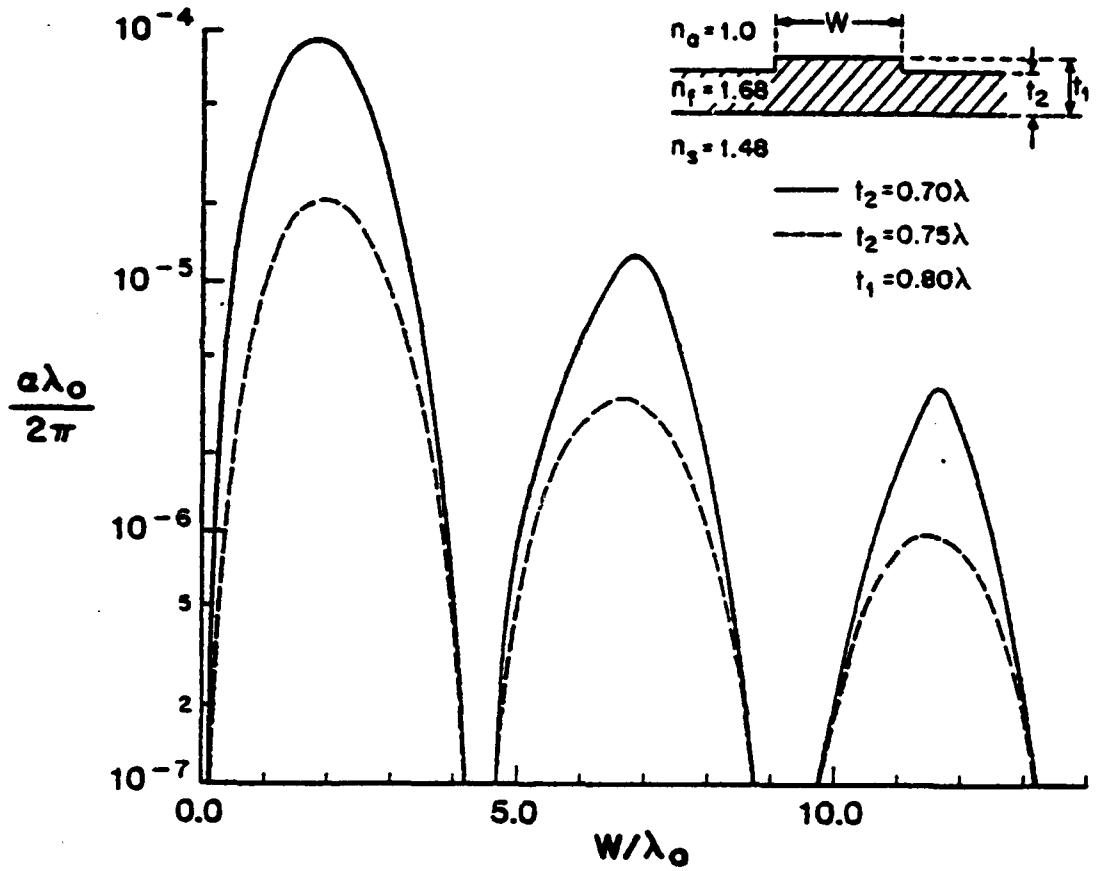


Fig. 3 Plot of attenuation constant α as a function of guide width W for a rib waveguide for integrated optics. Note the periodic sharp resonance dips.

In most cases, that wave will be propagating; when it is, as shown in Fig. 2, it indicates that energy in the TM polarization will be leaking per unit length along the waveguide, with the result that the propagation constant is complex, with a non-zero attenuation constant α.

It is important to note that the dominant portion of the guided wave energy possesses TE polarization but that the leakage energy has TM polarization, in the form of a TM surface wave propagating away from the waveguide at some angle. In all other known cases of leaky modes, the leakage energy possesses the same polarization as the exciting energy. The class of leaky modes described here thus constitutes a new class of such modes.

A typical plot of the attenuation constant α of the leaky mode as a function of the guide width W is shown in Figure 3. For a purely bound mode, the value of α would be identically zero (since we are neglecting material losses here). The non-zero value of α is thus due directly to the TM wave outside in Figure 2. We observe that the curve in Fig. 3 also shows periodic large dips, which are "resonances" or cancellation effects. These strong dips are due to the mode-converted TM wave in the inside region of Fig. 2, which also bounces back and forth between the sides of the waveguide. Thus, the mode-converted transmitted wave produces the leakage, and the mode-converted reflected wave produces the resonance dips in Figure 3. These are severe physical effects, not minor ones, and they are completely unexpected on the basis of the earlier published theories, which neglected the mode-conversion effects.

It is also worth noting that a recent publication[28] presented an analysis that took into account the TE-TM mode coupling but nevertheless did not recognize that leaky modes could be present. These authors were not sensitive to possible new physical effects and did not look for complex roots, thereby missing completely the only important new aspects of their analysis. In contrast, we recognized that the physical process of leakage must be present even before we conducted an accurate mode-matching analysis.

We have made calculations on most of the waveguides shown in Figure 1. We have shown on fundamental grounds[11,12] why the dielectric image guide for millimeter waves, and guides (a) and (b) for integrated optics, never leak, and why most modes on the remaining waveguides will ordinarily leak. We have also found that the magnitude of the leakage may be large or small, depending on the type of waveguide, and we have shown why. We have also presented[11,12] specific detailed examples for several of the waveguides shown in Figure 1. For example, the leakage is generally small for waveguides for integrated optics, and for the inverted strip waveguide for millimeter waves, whereas it can become very large under appropriate circumstances for the insular guide. For a special case of the insular guide, the attenuation was as large as 4 dB/λ!

On those waveguides which can leak, one can design the dimensional parameters so that the leaky mode leaks for all values of strip width W or only for narrow strip widths. Since the accurate calculations for the leakage constant α are rather complicated, it would be valuable to have available a simple criterion or method by which one could predict

for what dimensions leakage would occur. A simple analytical criterion was developed first,[7] and more recently it was extended to encompass a general, essentially graphical, method that is based upon the easily-found dispersion curves for the separate TE and TM surface waves, inside and outside the strip region, that are the constituent elements of the final guided wave. A paper[8] on this simple and systematic method has been prepared and will be submitted for publication.

Since the waveguides in Fig. 1 are intended for application in an integrated circuit context, whether optical or millimeter wave, it is clear that leakage from these waveguides could cause cross talk between neighboring circuits, thereby affecting device performance. This feature could be particularly distressing if the leakage is both large and unexpected. On the other hand, it should be possible to design new devices based on the new physical effect. As an illustration, we proposed a novel leaky-wave directional coupler[9] that could also serve as a mode stripper or mode purifier. That proposal was for integrated optics. We have since recognized that a modified version could be useful at millimeter waves, and we explored its feasibility during the past year.

A model similar to the one for integrated optics[9] could be designed in direct analogy, but it would also leak away from both sides of the coupling region. A better version would be a coupler that does not leak into the outside regions, but only between the two parallel waveguides that comprise the coupler. The usual type of directional coupler also consists of two parallel strip waveguides, but with the spacing between them sufficiently small that one waveguide is in the exponentially decaying transverse field of the other. In the leaky-wave coupler, the spacing between the parallel strip waveguides is so large that the coupling between the guides due to the decaying transverse fields is negligible; the coupling between them can then be due only to the leaky wave existing in the region between the strips. The difficulty referred to above is that, in the original design, when leakage occurred between the strips it also occurred outside of them.

The unwanted leakage outside of the strip coupling region can be eliminated by slightly raising the height of the coupling region between the strips; if the design is proper, each strip waveguide leaks only into the region on its side which is raised. The design worked satisfactorily with respect to the coupling desired, and it eliminated the unwanted outside leakage. However, the raised section between the strips unfortunately permitted the existence of additional bound modes, some of them occurring in just the range of directional coupler action. Rather than attempting to modify the structure further in order to eliminate these extra modes, we dropped the project, since we are concerned primarily with understanding the basic effects and not with device development. Some of the details relating to this investigation are included in a portion of a Ph.D. thesis.[29]

It was mentioned above that we have devised a simple and practical graphical procedure[8] to tell whether a particular structure will or will not leak, and also for what dimensions leakage will begin or stop. That method, however, does not indicate the amount of the leakage. For that it is necessary to solve the boundary value problem, which can be very complicated if very accurate values of leakage are desired. On

the other hand, for many applications it is sufficient to know the amount of leakage only approximately. We are currently exploring approximate methods towards that end.

Originally,[3] our numerical values for leakage were computed by including only the surface waves and their interactions, and neglecting the contributions from the continuous spectrum, which would change the numerical values somewhat but would not introduce any new physical effects or alter existing ones. A paper[30] by Japanese authors appeared in March 1979 which observed that our published results neglected the continuous spectrum; they then presented results based on an accurate, but still approximate method for taking into account all modal contributions and solving for the leakage. They compared their results with our admittedly approximate ones, and they concluded that ours were too small by an order of magnitude. In their paper, they criticized the usefulness of our approximate procedure because of the large discrepancy. However, they later informed us that they found a numerical error in their results, making them smaller by a factor of $(1.6\pi)^2$, which is about 25. They published a correction[31] to their numerical values, and presented a new curve which shows much closer agreement with our approximate values, but they unfortunately did not remove their earlier critical statement. In any case, when we compare our own accurate results to their published ones, we find very good agreement except in the vicinity of the resonance, or cancellation, points. Our results show very sharp dips, as seen in Fig. 3, whereas their dips are filled in substantially, because of the approximate method they used to compute the attenuation constant which yields the leakage.

Since our only real disagreement with the paper by the Japanese authors was the nature of the dips, we examined carefully the leakage behavior in the neighborhood of these dips. During the past year, we found the following interesting set of results. When we took into account only the constituent transverse surface waves in the solution, the dips were rigorously found to be nulls. We discretized the continuous spectrum contributions, and then took into account only one additional transverse mode, as an approximation to the continuous modes. That small modification was sufficient to soften the null and thereby produce only a sharp dip instead of a perfect null. Taking more transverse modes into account served to soften the null further, but the dips produced always remained very sharp. Accurate numerical calculations showed that the dips were about three orders of magnitude below the maximum in the curve of the leakage constant α as a function of the strip width or the frequency. We conclude, therefore, that the resonance dips in the leakage constant are not nulls, but that they nevertheless are very sharp and deep. Our results for the complex propagation constant (which of course includes the attenuation, or leakage, constant) are obtained by directly locating the complex roots. On the other hand, the Japanese authors first located the real part β of the complex root accurately, and then, using that value for the real part, obtained the imaginary part α by means of an accurate approximation. Their method, which is less direct and accurate than ours, was therefore able to yield very good results for α except in the neighborhood of the dips.

Our proof with respect to these dips proceeded in terms of a rigorous equivalent network comprised of ideal transformers representing

the coupling elements, and transmission lines for each of the surface waves and the discretized modes corresponding to the continuous spectrum. At each stage, i.e., taking into account only certain modes, the input impedance was examined; when modes corresponding to the continuous spectrum were included, the input impedance was found analytically to contain a resistive part that could not be cancelled out, so that the dips had to represent incomplete nulls. This study forms a portion of a recently completed Ph.D. thesis.[37]

We were gratified to note a recent publication[10] by Japanese authors which describes an experiment they performed to detect leakage from a rib-type dielectric waveguiding structure. They make strong reference to our theoretical work and our physical explanation for the effect, and they conclude that their measurements verify the theoretical predictions.

A basic difficulty, or limitation, associated with that equipment is that it is indirect. The leakage corresponds only to one constituent of the total insertion loss, and the measurement simply shows that the total insertion loss becomes modified in a manner consistent with our theoretical predictions when certain parameters are varied. We therefore set up a measurement program of our own in which the leaky wave and its properties would be measured directly.

Our direct measurement program encountered difficulty with spurious radiation effects. The building of the set-up was completed some time ago, but unexpected radiation contributions have distorted the measured results. It is clear that leakage does occur, and in the proper polarization; in fact, part of the problem is that the leakage is so strong that multiple reflections of that energy are found. Theory predicts that the energy should leak at a specified angle, and that, as one probes in a direction perpendicular to the guide axis, the leakage field should increase up to a critical distance and then drop off rapidly. We find that the angle agrees well quantitatively with theory, and that the transverse variation follows the theory reasonably well, except for interference effects which are quite strong. The primary source of the interference is a spurious space wave which emanates from the feed horn, and which seems very difficult to eliminate. These measurements do verify that the leakage obtained is indeed due to a leaky wave, is in the polarization expected, occurs at the right angle, and roughly corresponds to the expected transverse field variation. The strong interference, however, did not permit satisfactory quantitative correlation with theory for some of the parameters.

In connection with measurements of this type, we established a collaboration this past year with Professor H. Shigesawa and his colleagues at Doshisha University, Kyoto, Japan. Our measurements were taken at a wavelength of about 3 cm, whereas those in Japan were taken at a wavelength of 6 mm (f = 50 GHz), so that a greater waveguide length could be probed well. Spurious radiation also affected their measurements, but the greater length available relative to wavelength minimized the effects of the spurious radiation. In addition, those measurements in Japan demonstrated clearly that no radiation was measureable in the resonance regions where sharp dips in α were obtained theoretically. A joint paper that combined our measured results with theirs was presented recently at a symposium.[40]

During the past year, we devised a novel alternative measurement approach, based on a cavity resonance procedure. Suppose the original dielectric waveguide is comprised of a dielectric strip on a dielectric layer on a ground plane, and that the strip extends in the z direction. We then place a length a of that guide between parallel vertical metal plates, where the metal plates extend in the x and y directions, thus forming a resonant cavity. In that cavity, between the plates, we then have a strip of dielectric of width a extending along the x direction, with its center portion, of "length" W, having a somewhat greater height, since it corresponds to the strip of the original strip dielectric guide. If the mode on the original dielectric strip guide were purely bound, the fields in the x direction would be evanescent; if the mode were leaky, and a TE mode were incident, then a TM surface wave would leak away in the outside region.

The measurement approach employing the cavity now reverses the process. A TM wave is sent in from one end between the parallel planes, in the x direction, toward the central section of greater dielectric height; the frequency is then varied. If the condition for leakage were satisfied, the incident TM wave would excite a resonance in the central section, and the fields there would increase substantially, but only in the TE polarization (the inverse of the leakage situation). Hence, a probe sensitive to TE fields is placed just above the central section, and the power detected is found to peak sharply when the frequency of excitation corresponds to the condition for resonance. The fields scattered by such a cavity structure were also calculated theoretically, and the agreement with measurements was excellent. Measurements were taken over a variety of strip widths and frequencies, and those measurements were also compared with computations of the guide wavelengths from which the cavities were created. Very good agreement was found between the guide wavelengths and the resonance frequencies, on the one hand, and between the guide attenuation constants and the Q's of the cavities.

The novel cavity method verified that leakage indeed occurs, but it also offers a new method of measuring the leakage constant of these leaky waveguides. Resonant cavity methods have been used previously for the measurement of leakage constants of leaky modes, but the arrangement here is different. In other methods, a resonant cavity is created in the same way, by establishing metallic bounding planes, but there the guide section is excited longitudinally, via coupling holes in the metal planes. Here, the structure is excited transversely, by sending in the wave which would have leaked in the reversed situation. The details of this new measurement method, and the results of these measurements and their comparisons with theory, are contained in a Ph.D. thesis.[33]

It should finally be noted that, in recognition of the extent of our investigations and the importance of our contributions, we have been invited to present a pair of comprehensive papers[11,12] which keynote a special issue on open waveguiding structures.

C. Oblique Guidance By a Planar Periodically-Grooved Dielectric Layer

This study began as a first step in the rigorous analysis of radiation from periodically-grooved grating antennas on millimeter-wave

dielectric antennas, but we soon realized that the solution is useful and applicable in its own right in integrated optics. The study would then be that of optical surface waves propagating at an angle on a periodically-corrugated dielectric layer on a dielectric substrate. Such structures, with obliquely-incident surface waves, have been finding increasing application lately in such components as multiplexers, filters and mode deflectors.

The guidance of optical surface waves propagating in a direction normal to the grooves on a periodically-corrugated planar dielectric waveguide is well known, and various theoretical treatments are available, some of the most accurate having been developed at our Institute. Such structures have been used in grating couplers and in distributed-feedback lasers, for example. The electromagnetic boundary-value problem for the case of normal incidence is scalar, however, where the TE and TM surface waves remain independent of each other. When the surface propagates at an oblique angle with respect to the grooves, the TE and TM surface waves are now coupled together, and the boundary-value problem becomes a three-dimensional vector one. The problem is one for which an exact solution has previously not been available.

We have during this period derived a rigorous analytical solution to the vector, oblique-incidence problem which is valid for all ranges of parameter values; we believe that this solution represents the first exact analysis of a three-dimensional vector guided-wave problem involving a periodic structure.

A rich variety of physical effects arise as a result of oblique guidance, which are not present at all for normal guidance. One immediate consequence, due to the TE-TM coupling which occurs at every groove step, is that of polarization conversion to the other mode type; that is, if a TE mode is incident, some energy would be mode-converted into a TM mode. Another effect, explained below, is that at stop bands, corresponding to Bragg reflection, additional stop bands appear. Although many other interesting effects can arise, these two effects were recognized previously in two papers[34,35] which also presented approximate theoretical treatments valid only for small periodic perturbations, and only when the waves are purely bound. Our rigorous analysis is of course valid for rectangular grooves of arbitrary depth, aspect ratio or period.

Suppose we consider a grooved planar dielectric wageguide with the grooves parallel to the y direction, and the x direction perpendicular to them. The propagation behavior of waves guided at an oblique angle to these grooves is represented by a wavenumber plot, β_y vs. β_x. Such a plot indicates the behavior as a function of angle with respect to the grooves, but at a fixed frequency. Alternatively, a k vs. β plot presents the behavior as a function of frequency for a given angle. In both of these plots, the periodicity in the x direction, which causes space harmonics to be excited, produces additional curves on these plots repetitive in x. Where certain of these curves cross, stop bands are present, the first of these stop bands being known also as Bragg reflection.

Separate curves are present for each possible mode, and then for each space harmonic of each mode. Usually, the waveguide thickness is

such that only one TE surface wave mode and one TM surface wave mode are present. If we then consider the Bragg reflection condition, that stop band corresponds to the crossing of the n=0 (basic) wave and its n=-1 space harmonic. When no coupling exists between the TE and TM surface waves, we obtain one TE and one TM Bragg stop band. However, when oblique guidance occurs, the TE and TM surface waves do couple, and four stop bands are possible instead of only two. The two additional stop bands correspond respectively to the n=0 TE and n=-1 TM curves, and to the n=0 TM and n=-1 TE curves, couplings which could not occur for normal guidance.

We were fortunate in learning about the availability of accurate experimental data in the above-mentioned stop-band regions. We first learned of these measurements at a Workshop,[36] and then further at an Optical Society Topical Meeting.[14] The speaker, Dr. Reinhard Ulrich of Germany, indicated that no accurate theoretical data were available with which he could compare his measurements, and he was delighted to learn of our theory. Dr. Ulrich provided us with detailed data, and we developed a computer program for the required numerical values.

A comparison between our theory and his measurements showed excellent qualitative agreement, with stop bands wide where they should be, etc.; that comparison was shown in last year's proposal as Fig. 4. We had understood that the grooves on which the measurements were made were more sinusoidal than rectangular; since our theory applies to rectangular grooves, we did not expect that the comparison would be numerically exact. In September 1982, however, on a side trip associated with attendance at the European Microwave Conference, Professor Oliner visited Professor Ulrich in Hamburg, and spent two days with him discussing these measurements and planning future collaboration. It turned out that the dimension given to us earlier by letter for the dielectric layer thickness was the value before the grooves were etched. The etching process also reduced the layer thickness somewhat; although the final thickness was not known, a good estimate was available for the probable range of the reduction in thickness. In addition, we learned from the Ph.D. student, R. Zengerle, who was familiar with the intimate details, that an SEM photo of the structure had shown that the grooves were primarily rectangular, not sinusoidal as we were informed initially.

We therefore recalculated our theoretical results by employing a more correct value for the layer thickness. We took the middle of the probable range for the reduction, which then reduced the layer thickness from 0.165 μm to 0.159 μm. This small change produced a dramatic result. The theoretical values now agree quantitatively with the measurements to a remarkable degree. This new comparison is shown in Fig. 4, which replaces the earlier Fig. 4. It should be noted that the region shown in Fig. 4 is an enlargement of the area of the wavenumber plot just in the neighborhood of the stop bands.

Part of the planning mentioned above for future collaboration with Professor Ulrich involved a series of calculations for us to do, which he believed would reveal some additional physics in the stop band behavior, as well as be of value to designers of optical multiplexers or filters. The aim was to determine the variation of the widths of the three different types of stop bands (TE-TE, TM-TM, and TE-TM) as a function

of various geometrical parameters, such as groove depth while maintaining groove thickness constant, etc. When we made such calculations, we found some interesting behavior, which we were able to understand and explain fully. These new calculations and their implications, together with the new agreement shown here as Fig. 4, were presented jointly with Ulrich at a symposium[41] in June 1983.

Various physical implications follow from these wavenumber curves. First, we recognize that in this type of plot the phase-velocity direction is given by the line from the origin, whereas the group-velocity or energy-flow direction is given by the perpendicular to the actual curve. Thus, in the vicinity of the stop bands there is substantial anisotropy; that is, the energy-flow direction can be quite different from the direction of phase progression. This effect is called beam steering.

An illustration of beam steering is presented in Fig. 5, where for simplicity only one of the mode types is included. Two different directions of propagation are shown; for the one which is tilted further away from the x axis, the directions of phase progression and energy flow are about the same, but, for the other one, substantial beam steering occurs. For the ray which corresponds to strong beam steering, it is seen that a small change in the incident direction may result in a large change in the energy direction. Similarly, if the frequency is changed, resulting in a vertical shift in the dispersion curve, a large change in energy direction may be produced even if the direction of incidence is kept the same.

The different energy-flow directions shown in Fig. 5 can be utilized to produce focusing. What is interesting here is the unusual physical situation in which beam focusing can be achieved by a grating with only straight-line geometry.

Another set of interesting physical effects occurs when the frequency is raised sufficiently for one or more of the space harmonics to become radiating. The first space harmonic to radiate is the n = -1 space harmonic, and it will radiate first into the substrate. For a given incidence angle into the periodic grating, one can determine how many beams will radiate and where. If one examines carefully the exit angle of the n = -1 space harmonic, he finds very peculiar results. In general, the beam radiates at some skew angle with respect to the plane of incidence. We have not as yet examined this skew effect carefully, but we can see that the potential exists for unusual opportunities.

An additional new physical effect that arises because of the TE-TM mode coupling produced by oblique guidance is cross polarization. Again, this effect is absent entirely when the guidance is perpendicular to the grooves. The amount of cross polarization cannot be assessed without employing the complete dynamic theory, which we have done during this past year.

During the past year, therefore, we have extended the basic theory to encompass this new direction. Our previous results permitted us to compute the propagation characteristics of these surface waves guided obliquely on a planar grooved dielectric waveguide for integrated optics. Such computations led to dispersion curves for the behavior of these waves. Mathematically phrased, the solutions obtained yielded the

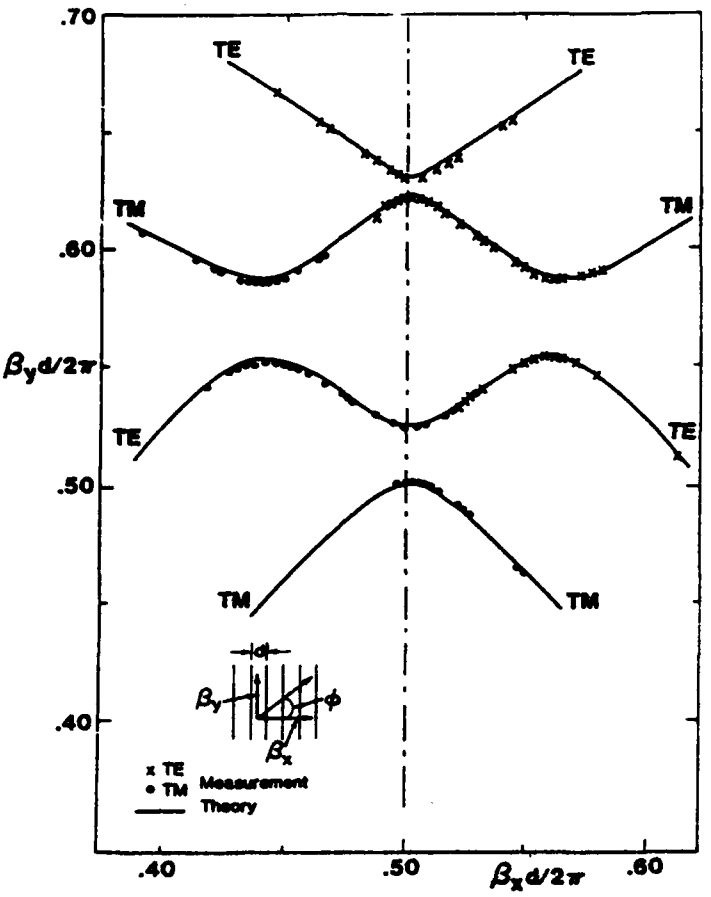


Fig. 4 Bragg interaction region for periodic optical waveguide. Comparison between rigorous theoretical calculations (solid line) and accurate measurements (points) taken by Ulrich and Zengerle[14] for the stop band region of the wavenumber diagram for surface waves propagating at an oblique angle ϕ to a periodically grooved dielectric layer on a dielectric substrate. Note that the ordinate and abscissa scales are greatly expanded, corresponding to enlargement of the wavenumber diagram just in the neighborhood of the stop bands. The top and bottom stop bands, centered at $\beta_x d/2\pi = 0.50$, are due to TE-TE and TM-TM interactions; the TE-TM stop bands are shifted from the customary 0.50 value. The various parameters are: groove thickness $t_g = 0.048$ μm, film plus groove (total layer) thickness $t_f + t_g = 0.159$ μm, period $d = 0.282$ μm, free space wavelength $\lambda_o = 0.6117$ μm, refractive indices of film n_f and substrate n_s are, respectively, 2.10 and 1.47.

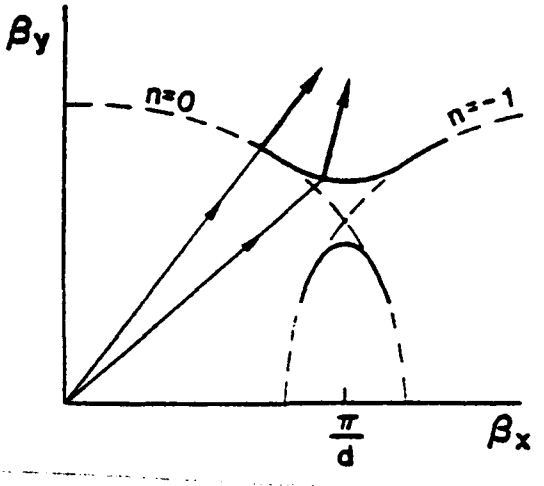


Fig. 5 Beam steering near a stop band.

eigenvalues for these guided surface waves. The theory was now extended to include the eigenvectors, in mathematical phrasing; that is, the solutions now also include the relative amplitudes of all the field components. The theory now permits us to determine the polarization content of the complicated fields that result when the TE and the TM constitutent surface waves become coupled at each step of the grating.

We have applied this extension of the theory to the case of groove spacing wide enough to permit the n=-1 space harmonic to become radiating; all the other space harmonics remain bound, however. When the surface wave is guided in the direction normal to the grooves, the polarization of the radiating space harmonic remains simple, the same as that of the incident surface wave. As that surface wave is rotated with respect to the grooves, corresponding to oblique incidence, the polarization content becomes mixed.

We expected a significant amount of cross polarization to be present, but we were actually surprised by the strength of it. For normal incidence on the grooves there is no cross polarization, of course, but the amount becomes quite strong even for small deviations from normal. For an incident TE surface wave, for example, the TM content of the radiated power actually exceeds the TE content for angles from the normal greater than 15° or so. For an incident TM surface wave, even stronger cross polarization effects are found; for example, in one case, for an angle somewhat less than 15°, a maximum occurs at which about 90% of the radiated power consists of the "wrong" polarization. Some of these results were presented at a symposium[42] in May 1983, and these studies of polarization content also form part of a recent Ph.D. thesis.[37]

The discussion above has all been concerned with the guided wave behavior of surface waves on periodically grooved dielectric layers. One should note, however, that the derivation of such guided wave theoretical

results involves a transverse resonance procedure, or, equivalently, a self-sufficiency condition in the transverse plane. The self-sufficiency condition here involves the multimode scattering of plane waves incident at an angle on the periodic grooves. Thus, a constituent part of the guided wave problem is the subsidiary problem of the scattering produced by a plane wave of arbitrary polarization content incident at an arbitrary angle in both orthogonal directions on a periodically grooved dielectric layer. That subsidiary problem was investigated independently during this past year, and some new results were found, including resonance effects and polarization interchanges. These results were presented at a symposium[43] in August 1983, and some of the new features were included in a recently completed Ph.D. thesis.[37]

D. Dielectric Gratings of Finite Width, and Implications for Device Performance

Dielectric waveguides for millimeter-wave integrated circuits possess narrow widths so as to permit single-mode operation. Many people have proposed that gratings be placed on such waveguides to achieve two types of device: resonators and filters, and leaky-wave antennas. For the resonators and filters, the gratings are operated in the Bragg reflection region so as to achieve strong reflections combined with negligible radiation. For the antenna application, the gratings are operated in the leaky-wave region so as to achieve controlled radiation with negligible reflection.

Since wave guidance by these open dielectric waveguides may be viewed in terms of a pair of surface waves propagating at an angle to the waveguide axial direction, the new physical effects discussed under C above apply here as well when a grating is placed on the waveguide. This situation is illustrated in Fig. 6 when a grating is placed on a dielectric image guide. Some differences are present, of course, between this situation and that of oblique guidance on an infinitely-wide grating, in that here a pair of waves, rather than a single wave, must be considered, and the influence of end effects must be included.

Our preliminary investigation of the effect of finite width on resonator or filter performance has shown that when the grating of finite width is operated in the Bragg regime an extra (unwanted) transmission dip or reflection peak may be obtained, which may seriously impair the device performance. The extra reflection peak near to the desired peak arises from an extra stop band due to interaction between TE and TM space harmonics, and in fact has a polarization opposite to that in the desired peak. This effect thus corresponds to the presence of the extra stop bands shown in Fig. 4, which were due to such TE-TM coupling. No such problem was encountered with SAW resonators, where the grooves are very wide, and it was assumed that narrowing the width would not introduce any new problems. That supposition can be quite wrong because a new physical effect is introduced as a result of the finite width, as we pointed out recently.[15]

It is worth noting that Japanese authors[35] have recently devised a multiplexer for integrated optics that utilizes two periodic gratings at an angle to the incident surface wave. They found experimentally that an extra reflection peak, of opposite polarization, arises in this process due to the TE-TM coupling.

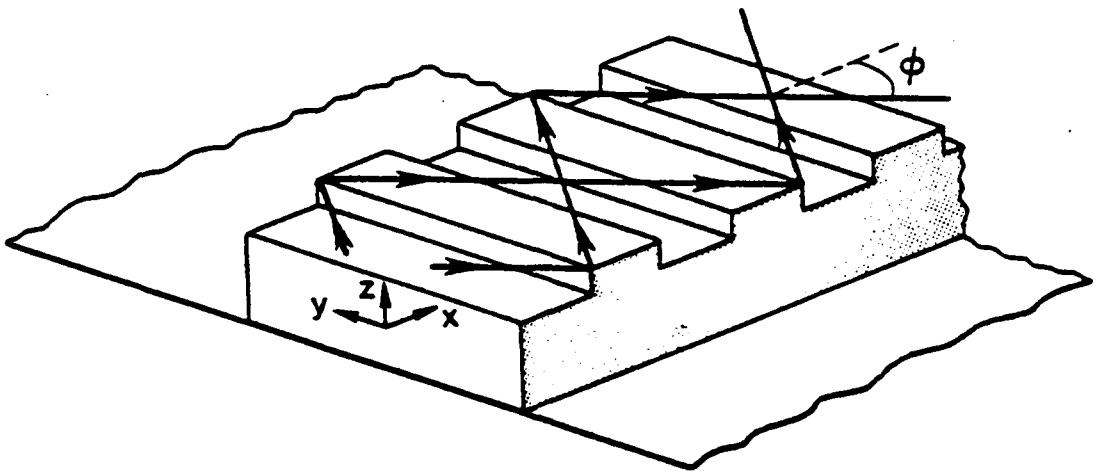


Fig. 6 The guidance of surface waves along a periodically-grooved dielectric image guide, showing constituent waves propagating at an angle.

Many measurements have been taken on leaky- wave antennas comprised of gratings on dielectric image guide, of the type shown in Fig. 6. No theory for the leaky-wave behavior had appeared, however, taking finite width into account, until our recent presentations.[18,19] Our theory adapts the analytical result mentioned earlier for oblique guidance by a grating and presents a straightforward procedure for determining the α and β of the leaky wave, thus permitting one to design such an antenna in accordance with desired radiation characteristics. Ours is the only theory available to date that yields such information.

That design suffers in only one way, however. It is based on the rigorous theory for oblique guidance by a grooved dielectric surface, which is rather involved mathematically and therefore has an associated complex computer program. A practical design procedure should require a less involved computer program. Toward that end, we derived during the past year an alternative and simpler procedure, based on a perturbation analysis that yields simple closed-form solutions for the radiation characteristics of these periodic structures.

In this new analysis, the radiation of the surface wave in the presence of the periodic corrugation is formulated in terms of the radiation from an equivalent distributed dipole source. That source is due to the interaction of the surface wave with the periodic corrugation which is taken as a perturbation on a uniform (unperturbed) multilayer dielectric structure. Such a method was originally developed by us for the special case of normal incidence, where the TE and TM polarized waves exist independently; many useful closed-form solutions were then obtained for the two polarizations separately. We have observed that the equivalent dipole source vector for the general case of oblique

incidence may be decomposed into two orthogonal vectors, each responsible for the radiation of one polarized wave. Thus, the closed-form solutions previously obtained for the normal-incidence case can be simply employed for the oblique-incidence case, with only a slight modification of the TE and TM source amplitudes that account for the important effect of cross coupling between the two polarizations.

For best perturbation results, the propagation wavenumber β of the "unperturbed" surface wave must be chosen properly. It turns out that for TE mode incidence (when $H_z \neq 0$ and $E_z = 0$), the value of β for the surface wave should be based on the volume average of the dielectric constant ε, taking into account the tooth and gap regions of the grating. For TM mode incidence (when $E_z \neq 0$ and $H_z = 0$), the β value should be obtained from the volume average of the refractive index n ($\varepsilon = n^2$).

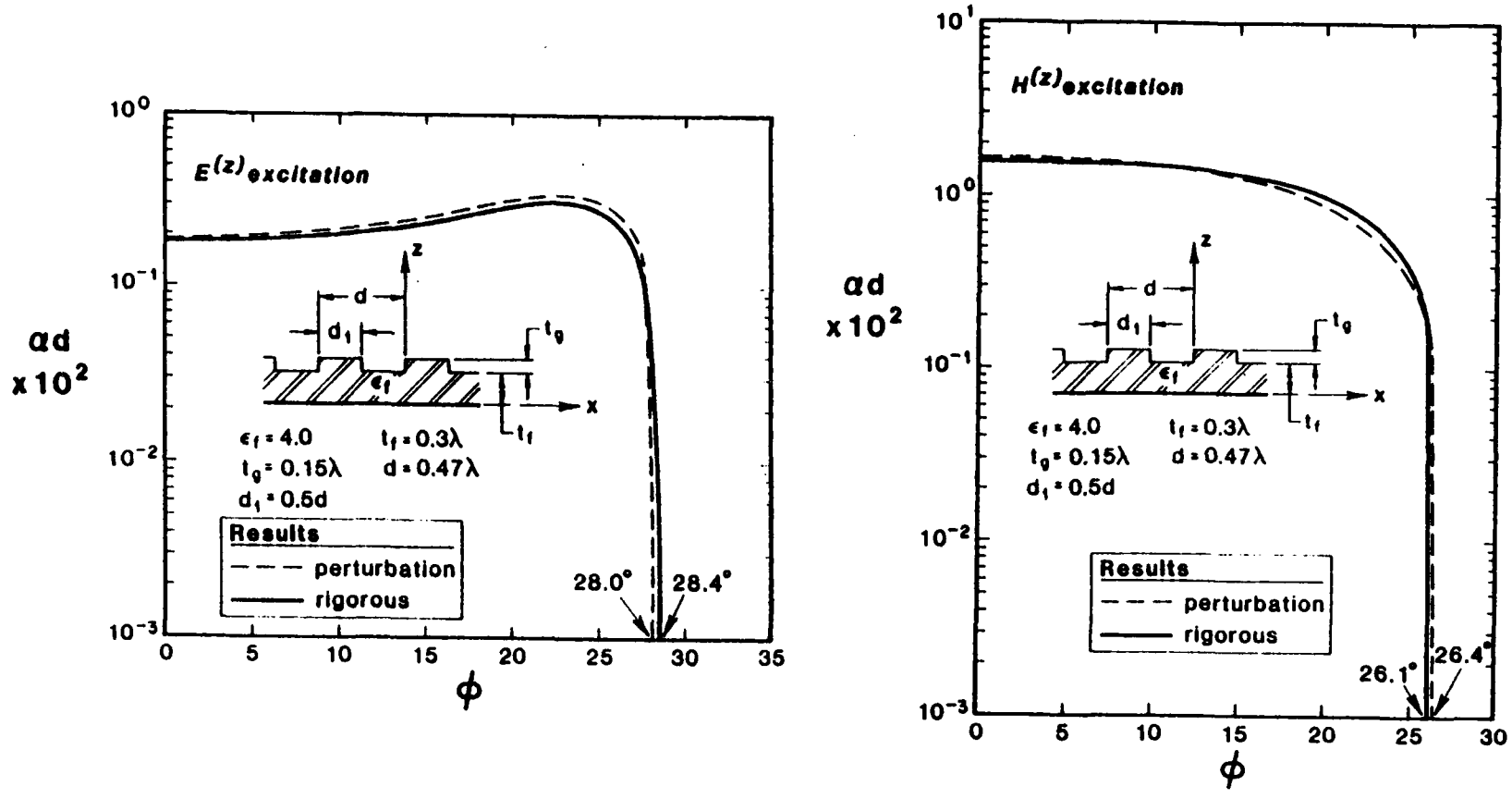
In order to check the accuracy of the perturbation expressions, a typical specific antenna geometry was selected and the numerical results obtained were compared with those found using the rigorous expressions. The value of the attenuation constant α due to leakage was computed as a function of angle ϕ, where ϕ is shown in Fig. 6. The value of angle ϕ depends on the guide width, being larger for narrower guides. These values of α were computed at a specified frequency, but for both polarization excitations.

The first comparison, shown in Fig. 7(a), is made for vertical incident electric field polarization, corresponding to the lowest TM-like mode on a heavily-grooved image guide, which is the dominant mode on that guide. The solid line in Fig. 7(a) corresponds to the accurate but complex theory. Calculations derived from the new perturbation analysis are indicated by the dashed line. The agreement is seen to be very good, despite the simplicity of the perturbation expressions. In Fig. 7(b) the comparison is presented for horizontal incident electrical field polarization, corresponding to the lowest TE-like mode on the structure. Again, the solid and dashed lines represent the accurate and approximate results, respectively, and again the agreement is seen to be very good.

The very good agreement obtained between the solid and dashed curves in Fig. 7 indicates that the perturbation analysis yields rather accurate results. Since the resulting expressions for α are in closed form and are relatively simple, this approximate procedure should also be a practical one for the design of this class of dielectric grating antennas for millimeter waves.

A summary of the theory underlying the new perturbation procedure was presented recently,[38] as was its application to radiation from dielectric grating antennas of finite width.[39] Additional details are contained in a Ph.D. thesis.[37]

One physical effect which should be present on grating antennas of finite width is that of cross polarization, due to the TE-TM coupling produced at every groove step in the grating. So far, however, no measurements made on such antennas have tried to determine the possible cross polarization that may be present, so that no experimental



(a) TM Mode (b) TE Mode

Fig. 7 The attenuation constant α as a function of guide width (in the form of equivalent angle ϕ) for the lowest TM and TE modes on a heavily grooved dielectric image guide.

indications are available yet. On the other hand, we have found during this past year, as discussed under C above, that the cross polarization effects that arise due to the TE-TM coupling associated with oblique guidance can be surprisingly large. On that basis, we expect that strong cross polarization will be found at certain skew angles of radiation. We plan to examine quantitatively how these effects will manifest themselves in practical antenna situations. In addition, in an angular range near broadside, leaky- wave stop bands occur, and our early results[17] indicate that cross-polarization effects should be significant there.

When a number of these new physical effects are placed together, they form a cohesive and very interesting package. We therefore prepared an overview paper which we presented at a symposium[44] in August 1983. The presentation was very well received, with many compliments. Included in the presentation were the extra stop bands in the Bragg region due to TE-TM mode coupling, the comparison shown in Fig. 4, the beam steering and focusing possibilities near the stop bands, the variety of skew radiation angles possible, and some results on strong cross polarization.

The investigations so far have concentrated on effects involving dielectric strip waveguides and periodically grooved dielectric layers with obliquely incident waves, in which the grooves occur in only one direction. Although such investigations are not complete, and we are of course continuing those studies, we have also begun recently an examination of a somewhat more complicated structure: a dielectric layer grooved periodically in two orthogonal directions, resembling a waffle iron. The reason for the new investigation is that our earlier brief look into the behavior of such a structure indicated that an enormous variety of weird-looking wavenumber diagrams may be possible. The implication is that additional new physical effects may emerge from such a study. That look was performed "kinematically", i.e., by simply examining various possibilities graphically, employing the various unperturbed interacting wavenumber curves. Recently, we have begun an analytical investigation, taking a model in which the variations in one direction are independent of those in the other orthogonal direction. We have only just begun, but we already see that some of the crossings that we assumed earlier to be interacting are in fact not; certain additional wavenumber conditions must be satisfied before the interactions are physically real ones. There remains, however, a large number of interesting-looking diagrams that clearly warrant a further look into the behavior of such structures.

5. REFERENCES AND PUBLICATIONS

1. J.P. Hsu, S.T. Peng and A.A. Oliner, "Scattering by Dielectric Step Discontinuities for Obliquely Incident Surface Waves," Digest of URSI Meeting, p. 46, College Park, Maryland (May 1978).

2. S.T. Peng and A.A. Oliner, "An Improved Calculation Procedure for Scattering by Dielectric Steps," Proc. Internat. URSI Sympos. on Electromagnetic Waves, Paper No. 321B (August 26-29, 1980).

3. S.T. Peng and A.A. Oliner, "Leakage and Resonance Effects on Strip Waveguides for Integrated Optics," Trans. Inst. of Electronics and Communication Engineers of Japan, Special Issue on Integrated Optics and Optical Fiber Communications, Vol. E61, pp. 151-154 (March 1978).

4. A.A. Oliner, S.T. Peng and J.P. Hsu, "New Propagation Effects for the Inverted Strip Dielectric Waveguide for Millimeter Waves," International Microwave Symp. Digest, pp. 408-410, Ottawa, Canada (June 1978).

5. A.A. Oliner and S.T. Peng, "A New Class of Leaky Modes on Open Dielectric Waveguides," Digest Internat. Microwave Symposium, pp. 569-571, Orlando, Florida (April 30-May 2, 1979).

6. S.T. Peng and A.A. Oliner, "Leaky Modes on Waveguides for Integrated Optics," Proc. Vol. 239 of S.P.I.E., Guided-Wave Optical and Surface Acoustic Wave Devices, Systems and Applications, Paper No. 239-09, San Diego, California (July 28-August 1, 1980).

7. A.A. Oliner and S.T. Peng, "A New Class of Leaky Modes on Dielectric Waveguides for Millimeter Waves," Proc. Internat. URSI Sympos. on Electromagnetic Waves, paper 311B, Munich, Germany (August 26-29, 1980).

8. A.A. Oliner and S.T. Peng, "A Simple Graphical Procedure for Predicting When Modes Will Leak on Dielectric Strip Waveguides," to be submitted for publication.

9. E.W. Hu, S.T. Peng and A.A. Oliner, "A Novel Leaky-Wave Strip Waveguide Directional Coupler," Topical Meeting on Integrated and Guided Wave Optics, Paper No. WD2, Salt Lake City, Utah (January 1978).

10. K. Ogusu and I. Tanaka, "Optical Strip Waveguide: An Experiment," Appl. Opt., vol. 19, pp. 3322-3325 (1 October 1980).

11. S.T. Peng and A.A. Oliner, "Guidance and Leakage Properties of a Class of Open Dielectric Waveguides, Part I: Mathematical Formulations," IEEE Trans. Microwave Theory Tech. Vol. MTT-29 (Special Issue on Open Guided Wave Structures), to be published in September 1981. Invited Paper.

12. A.A. Oliner, S.T. Peng, T.I. Hsu and A. Sanchez, "Guidance and Leakage Properties of a Class of Open Dielectric Waveguides, Part II: New Physical Effects," same as reference 11. Invited Paper.

13. S.T. Peng, "Oblique Guidance of Surface Waves on Corrugated Dielectric Layers," Proc. Internat. URSI Sympos. on Electromagnetic Waves, Paper No. 341B, Munich, Germany (August 26-29, 1980).

14. R. Ulrich and R. Zengerle, "Optical Bloch Waves in Periodic Planar Waveguides," Topical Meeting on Integrated and Guided Wave Optics, Incline Village, Nevada (January 28-30, 1980).

15. M.C. Shiau, H. Shigesawa, S.T. Peng and A.A. Oliner, "Mode Conversion Effects in Bragg Reflection from Periodic Grooves in Rectangular Dielectric Image Guide," Digest International Microwave Sympos., pp. 14-16, Los Angeles, California (June 15-17, 1981).

16. A.A. Oliner and S.T. Peng, "New Physical Effects Due to Mode Coupling in Various Dielectric Structures," Goubau Memorial Session of URSI National Radio Science Meeting, Los Angeles, California (June 15-19, 1981). Invited Talk.

17. A.A. Oliner and S.T. Peng, "Mode Conversion Effects in Radiation from Periodically-Corrugated Dielectric Structures," Digest of URSI National Radio Science Meeting, p. 15, Los Angeles, California (June 15-19, 1981).

18. S.T. Peng, A.A. Oliner and F. Schwering, "Theory of Dielectric Grating Antennas of Finite Width," Proc. of IEEE AP-S Internat. Sympos., pp. 529-532, Los Angeles, California (June 16-19, 1981).

19. S.T. Peng and A.A. Oliner, "Radiation from Grating Antennas on Dielectric Waveguides of Finite Width," Proc. 11th European Microwave Conference, Paper No. B8.5, Amsterdam, The Netherlands (September 7-11, 1981).

20. H. Furuta, H. Noda and A. Ihaya, "Novel Optical Waveguide for Integrated Optics," Appl. Opt., Vol. 13, pp. 322-326 (February 1974).

21. V. Ramaswamy, "Strip-Loaded Film Waveguides," Bell System Tech. J., Vol. 53, pp. 697-704 (April 1974).

22. N. Uchida, "Optical Waveguide Loaded with High Refractive-Index Strip Film," Appl. Opt., Vol. 15, pp. 179-182 (January 1976).

23. W.V. McLevige, I. Itoh and R. Mittra, "New Waveguide Structures for Millimeter-Wave and Optical Integrated Circuits," IEEE Trans. Microwave Theory Tech., Vol. MTT-23, pp. 788-794 (October 1975).

24. T. Itoh, "Inverted Strip Dielectric Waveguide for Millimeter-Wave Integrated Circuits," IEEE Trans. Microwave Theory Tech., Vol. MTT-24, pp. 821-827 (November 1976).

25. S. Shindo and T. Itanami, "Low-Loss Rectangular Dielectric Image Line for Millimeter-Wave Integrated Circuits," IEEE Trans. Microwave Theory Tech., Vol. MTT-26, pp. 747-751 (October 1978).

26. E.A.J. Marcatili, "Dielectric Rectangular Waveguide and Directional Coupler for Integrated Optics," Bell System Tech. J., Vol. 48, pp. 2071-2102 (September 1969).

27. R.M. Knox and P.P. Toulios, "Integrated Circuits for the Millimeter Wave Through Optical Frequency Range," Proc. Sympos. on Submillimeter Waves, Polytechnic Press of Polytechnic Institute of Brooklyn, pp. 497-516 (April 1970).

28. R. Mittra, J.L. Hou and V. Jamnejad, "Analysis of Open Dielectric Waveguides Using Mode-Matching Technique and Variational Methods," IEEE Trans. Microwave Theory Tech., Vol. MTT-28, pp. 36-43 (January 1980).

29. J-S. Myung, "Guidance and Leakage by Open Dielectric Waveguides for Millimeter Waves," Ph.D. Thesis, Chapter V, Polytechnic Institute of New York (June 1982).

30. K. Ogusu, S. Kawakami and S. Nishida, "Optical Strip Waveguide: An Analysis," Appl. Opt., Vol. 18, pp. 908-914 (15 March 1979).

31. Ibid., "Correction": Vol. 18, p. 3725 (November 1979).

32. T-I. Hsu, "Propagation Characteristics of Dielectric Strip Waveguides," Ph.D. Thesis, Polytechnic Institute of New York (June 1983).

33. J-S. Myung, "Guidance and Leakage by Open Dielectric Waveguides for Millimeter Waves," Ph.D. Thesis, Chapter IV, Polytechnic Institute of New York (June 1982).

34. A. Gudzenko, "Bragg Reflection in Planar Dielectric Waveguides with Periodic Thickness Modulation," Radio Engineering and Electron Physics (USSR), Vol. 22, pp. 19-25 (1976).

35. K. Wagatsuma, H. Sakaki and S. Saito, "Mode Conversion and Optical Filtering of Obliquely Incident Waves in Corrugated Waveguide Filters," IEEE J. Quantum Electronics, Vol. QE-15, pp. 632-637 (July 1979).

36. Fourth Workshop on Optical Waveguide Theory, Noordwijkerhout, The Netherlands (September 1979).

37. M.J. Shiau, "Scattering and Guidance of Electromagnetic Waves by Periodic Dielectric Structures," Ph.D. Thesis, Polytechnic Institute of New York (June 1983).

38. S.T. Peng and M.J. Shiau, "Perturbation Analysis of Radiation from Periodically Corrugated Dielectric Layers: Oblique Guidance Case," Digest of National Radio Science Meeting, p. 20, Albuquerque, New Mexico (May 24-28, 1982).

39. M.J. Shiau, S.T. Peng and A.A. Oliner, "Simple and Accurate Perturbation Procedure for Millimeter Wave Dielectric Grating Antennas of Finite Width," Digest of IEEE International Symposium on Antennas and Propagation, pp. 648-651, Albuquerque, New Mexico (May 24-28, 1982).

40. H. Shigesawa, M. Tsuji, J.S. Myung, S.T. Peng and A.A. Oliner, "Direct Experimental Confirmation of New Leakage Effects on Open Dielectric Strip Waveguides," Digest of IEEE International Microwave Symposium, pp. 293-295, Boston, Mass. (May 31 - June 3, 1983).

41. S.T. Peng, A.A. Oliner, M.J. Shiau and R. Ulrich, "Design Considerations for Dielectric Bragg Reflectors for Oblique Surface Wave Incidence," Proc. Fourth International Conference on Integrated Optics and Optical Fiber Communication (IOOC), pp. 418-421, Tokyo, Japan (June 27 - July 2, 1983).

42. M.J. Shiau, S.T. Peng and A.A. Oliner, "Strong Polarization Conversion in Radiation from Surface Waves Incident Obliquely on a Grooved Dielectric Layer," Digest of National Radio Science Meeting, p. 103, Houston, Texas (May 23-26, 1983).

43. M.J. Shiau and S.T. Peng, "Scattering of Plane Waves by a Dielectric Grating Structure: General Incidence Case," Proc. URSI International Symposium on Electromagnetic Theory, pp. 175-178, Santiago de Compostela, Spain (August 23-26, 1983).

44. A.A. Oliner and S.T. Peng, "New Physical Effects on Periodically-Grooved Open Dielectric Waveguides," Proc. URSI International Symposium on Electromagnetic Theory, pp. 515-518, Santiago de Compostela, Spain (August 23-26, 1983).

6. DoD AND OTHER INTERACTIONS

(a) Professor Peng spent two summers (1977, 1978) at the US Army CORADCOM at Fort Monmouth, New Jersey, working with Dr. Felix Schwering on dielectric grating and taper antennas, under an LRCP arrangement with the US Army Research Office. These interactions have continued in the form of a post-LRCP contract and successor contracts.

(b) Both Professors Oliner and Peng, but particularly Professor Peng, have held discussions (since 1979) on electronically-scanned millimeter-wave dielectric grating antennas with Dr. Harold Jacobs of the US Army ERADCOM, Ft. Monmouth, New Jersey, and have supplied information or the results of analyses on several occasions.

(c) Professor Peng has been a consultant since 1981 on the topic of integrated optics for Dr. John Zavada of the US Army ARRADCOM, Dover, New Jersey.

(d) Professors Oliner and Peng participated in the millimeter-wave workshop sponsored by the US Army in October 1980 in Colorado; Professor Oliner presented a short talk on his JSEP-supported studies and was invited to be a member of the panel that evaluated the state of the art.

(e) Professor Oliner has a contract with the Air Force's RADC/ET (Hanscom Field) on novel millimeter-wave antennas, but other than the grating structures studied under JSEP sponsorship.

(f) A collaborative program has been established with Professor R. Ulrich of the University of Hamburg-Harburg, West Germany, on the topic of new physical effects due to oblique guidance on optical grating structures, where we do the theory and his group performs careful measurements. This collaboration resulted in a joint paper presented at the IOOC in Japan in June 1983.

(g) We have also established a cooperative program with Professor H. Shigesawa of Doshisha University, Kyoto, Japan, on experimental verification of the new leakage effects on dielectric strip waveguides that we predicted theoretically. This collaboration resulted in a joint paper presented at the IEEE International Microwave Symposium held in Boston in June 1983.

B. WAVE INTERACTIONS AND ABSORPTION RESONANCES ON OPEN LOSSY STRUCTURES

Professor T. Tamir

Unit EM3-2

1. OBJECTIVE(S)

To explore basic aspects of the scattering, guiding, and absorption of electromagnetic waves by stratified and/or periodic media having intrinsic losses. Unlike previously studied structures that were assumed to be ideally lossless, the present situations are characterized by physical parameters that include realistic absorption, conversion, or radiation losses. In particular, the investigation will focus on the recently observed phenomenon of anomalously high energy absorption that occurs under circumstances involving only slight losses.[1-9] For this purpose, the projected study will use simplified canonic configurations that characterize a wide variety of situations and thus facilitate the clarification of fundamental concepts. Specifically, the results obtained would be relevant to integrated-optical applications, to selective spectral filters for electromagnetic or optical purposes, to photo-detectors, to the guiding by and leakage from geological structures, as well as to analogous acoustic-wave situations involving propagation along planar interfaces.

2. APPROACH

The wave interactions and the resonant-absorption behavior of canonic configurations with realistic losses are studied by utilizing rigorous techniques[16,17] that have already been developed to a high degree of sophistication and success in the area of microwave engineering and in integrated optics.[15] In this context, extensive use is made of results and procedures already available from some of our earlier studies.[18-23]

In addition to well-established techniques, we use a novel approach that applies complex pole-zero analysis to the parameters describing scattering and guidance. As an illustration, the reflectance of stratified media is a well-defined known analytic function which fully describes the reflection, transmission, and guidance properties in terms of the poles and zeros located in the wavenumber plane. The tracking of the pertinent pole-zero loci in that plane can therefore serve both for the analysis and for the synthesis of structures having prescribed behavior. All of these methods are employed analytically, with subsequent verification by numerical computer methods[16] and possibly also by experimental observations.[9]

An important aspect of the program is that the pertinent wave interactions are studied by considering first the scattering of incident plane waves, which yield relatively simple but idealized descriptions of the pertinent process. However, the plane-wave results are then extended to the scattering of incident realistic (e.g., Gaussian) beams, which provide an accurate physical description of actual situations. The extension to the beam-wave case is carried out by using integrations in the complex wavenumber plane, which are evaluated by suitable asymptotic techniques and other analytical or numerical approaches.[17-23]

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

(a) We have applied the complex pole-zero analysis to examine the reflectance function that describes a lossy multilayered configuration. We have found that this analysis distinguishes between two varieties of such configurations, of which one is basically reflecting and the other is basically transmitting. In the absence of absorption losses in any of the media, the former configuration reflects all of the incident plane-wave energy. In contrast, the latter configuration transmits a substantial portion of that energy. When losses are introduced into any of the media, the analytical and physical behavior of the two configuration varieties are quite different and are best treated separately. We have discussed the pole-zero analysis in two meeting papers[25,26] and, in particular, we have shown how to apply it to design frequency (or angular incidence) filters by exploiting the leaky-wave properties of layered structures.

(b) For multilayered configurations of the basically reflecting variety, we have considered their absorption effects in the context of incident Gaussian beams. The results were presented in two meeting papers[26,28] and in a comprehensive journal article.[27] That article shows in detail, that, when a Gaussian beam is incident under critical conditions so as to couple strongly to a leaky wave supportable by the structure, the reflected field exhibits a lateral beam-shift behavior that is anomalous. In that case, the reflected-beam profile can be severely distorted and its shift may appear in a backward rather than the usual forward direction. Such a backward shift is quite unexpected and thus represents a new type of beam displacement in the context of multilayered media.

(c) We have performed a preliminary analysis of multilayered configurations of the basically transmitting type so as to obtain their behavior under conditions similar to those discussed in the foregoing item (b). A particularly interesting and novel aspect is that transmitting configurations can support two varieties of leaky waves. A first variety is similar to that which occurs in basically reflecting configurations. This first type of leaky waves accounts for radiation into both the upper and lower open regions. In contrast, the second variety radiates only in the denser of the two open (air and substrate) regions that bound the layered structure. The existence of the latter leaky-wave variety has been recognized only recently[24], but its significance had not been explored. We have therefore discussed the behavior of those waves at two meetings[29,30] and have pointed out that they may strongly affect the fields scattered by the pertinent structures.

(d) We have extended the analysis of Gaussian-beam incidence described in (b) above to configurations of the basically transmitting type. In this case, the multilayered structures account for a transmitted beam in addition to a reflected one. We have found that now both the reflected and transmitted beams are displaced and distorted, but their lateral shift and profile distortion are different from those occurring for beams in basically reflecting situations. A particularly

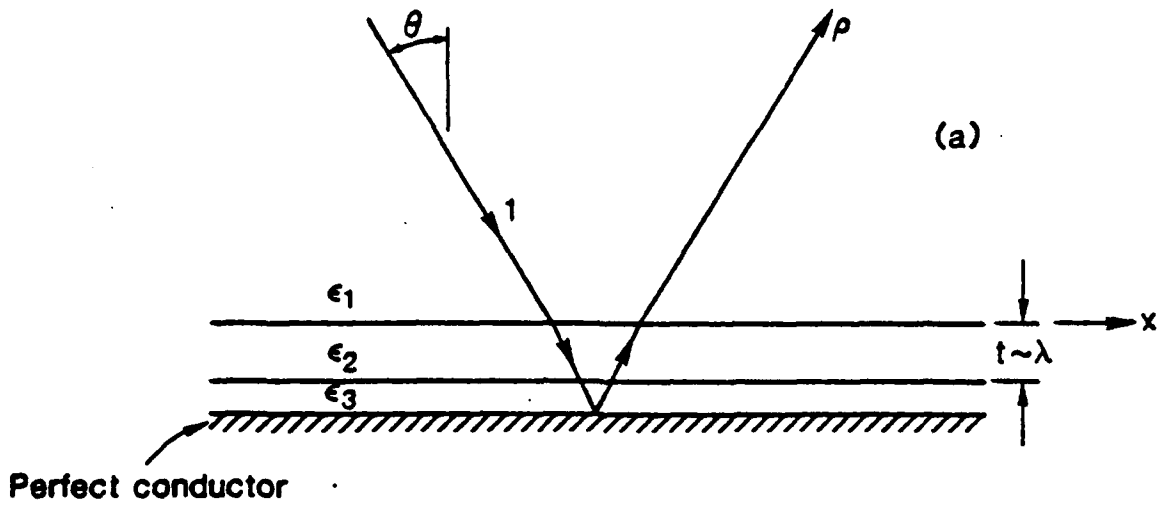
interesting aspect is that the transmitted beam always undergoes a forward shift. However, the reflected beam is shifted forward or backward depending on the given values of refractive indices of the layered configuration. This would permit interesting possibilities for controlling the shape and displacements of the scattered beams by appropriate choice of material constants. Most of these last results are going to be presented at a forthcoming meeting[31].

4. STATE OF THE ART AND PROGRESS DETAILS

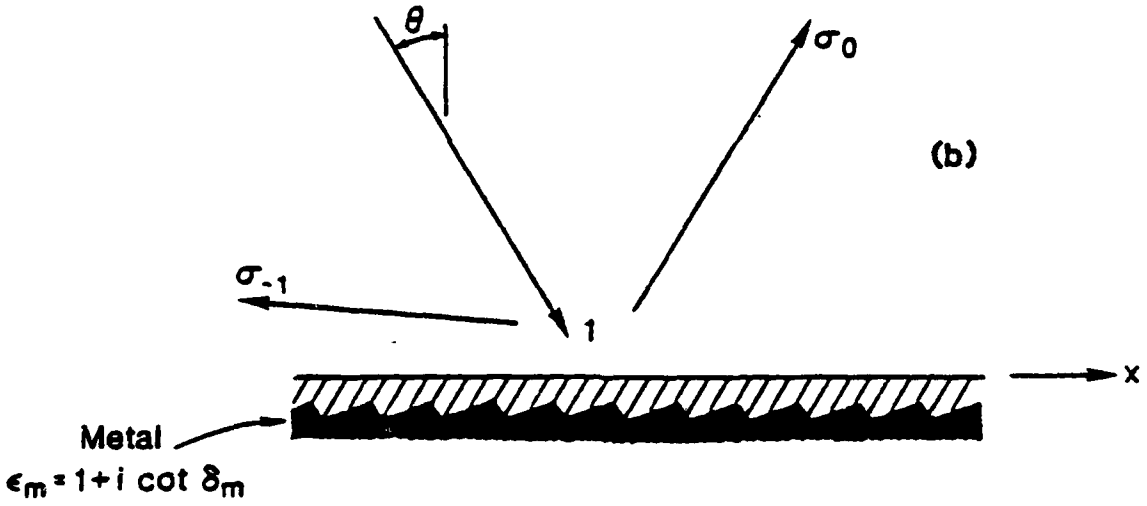
The propagation, scattering, and guiding of waves by planar configurations has been studied extensively in multilayered[10] and periodic[11] structures. However, most of these studies have examined ideally lossless configurations. The presence of losses usually has been treated as a perturbation on the ideal models, so that its effect was pronounced only if propagation occurred over very large distances. Thus, for small intrinsic losses, the scattered field is usually very similar to that of the lossless case, and the absorbed energy is then only a small percentage of the incident energy. As an example, for the simple case of a layered structure placed on a perfect conductor, as shown in Fig. 1(a), the reflectance ρ has a magnitude of unity. If small intrinsic losses occur in one of the layers, $|\rho|$ is generally close to unity for layers having a thickness t of the order of the wavelength λ, because the power flux then traverses a relatively small path inside the lossy medium, as suggested by the first-order refracted and reflected ray in Figure 1(a). A similar situation holds for periodic structures of the type shown in Figure 1(b). However, under critical combinations of the incidence angle and the physical parameters, the fields are strikingly different in that all of the incident energy is absorbed and the scattered waves are then suppressed.

This anomalous-absorption behavior was first identified in periodic structures[1,2,4] by the French group led by Professor Petit, but we have subsequently shown[3,7-9] that such a phenomenon occurs in a more general class of planar configurations, which may involve only a few layered media. A similar absorption behavior has been studied[12-14] in optics in the context of the so-called "induced transmission" through thin films. However, that transmission effect was restricted to normal incidence on stratified media, whereas the anomalous absorption phenomena to be explored occur at generally oblique incidence angles and they appear also in periodic structures.

Our recent studies have revealed[3,5,9] that, by taking a leaky-wave approach to the scattering phenomenon, it is possible to describe the anomalous effects in terms of a strong interactive coupling of energy between the incident wave and a leaky-wave field supportable by the structure. This interaction produces an electromagnetic-field regime whereby the energy flux travels over a lengthy longitudinal path D along the structure, as suggested in Figure 2. Thus, even for a layer having a thickness t of the order of the wavelength λ and very small losses per λ, the overall attenuation of the energy flowing along $D >> \lambda$ accounts for the anomalously high absorption effect. In this context, our previous studies have also demonstrated the important result that this effect can occur for both types of polarization,[3,7] and that it takes place in a wide class of multilayered structures.[9] In addition, we have experimentally verified[8] that the anomalous effect, which had previously been observed only in metallic gratings,[1,2] can nevertheless occur also in simple multilayered configurations.

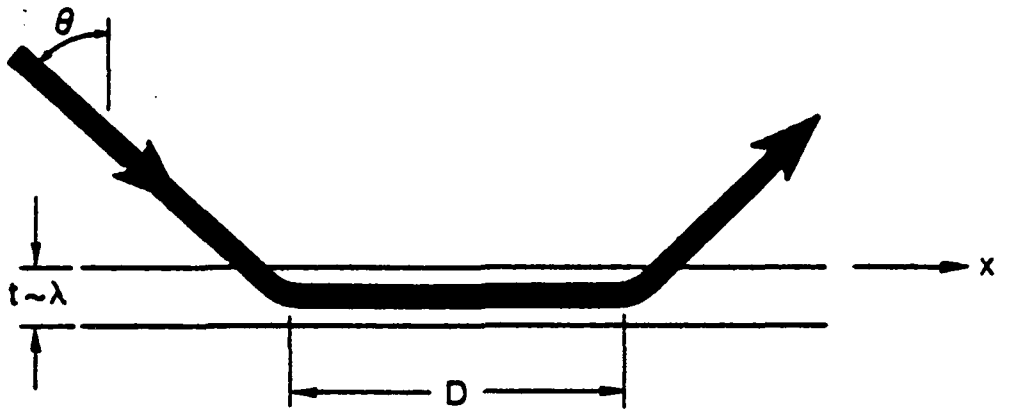


If some $\epsilon_j = \epsilon_j'(1+i\tan\delta)$: $|\rho| = 1-O(\delta)$



$$|\sigma_0|^2 + |\sigma_{-1}|^2 = 1-O(\delta_m)$$

Fig. 1 Scattering by thin planar structures. (a) Incidence and reflection at a layered configuration, (b) diffraction by a grating with only two propagating orders.



Leaky-wave field: $E \sim e^{ik_{xp}x} = e^{i(\beta_p + i\alpha_p)x}$

If $k \sin\theta = \beta_p$ and $\alpha_p << \beta_p$, then

$$D \simeq \frac{2}{\alpha_p} >> \lambda$$

Fig. 2 Leaky-wave interpretation of resonant absorption, showing how the incident energy is trapped into the structure in the form of a longitudinal power flux that flows along a distance D after which the energy is re-radiated back into the upper medium.

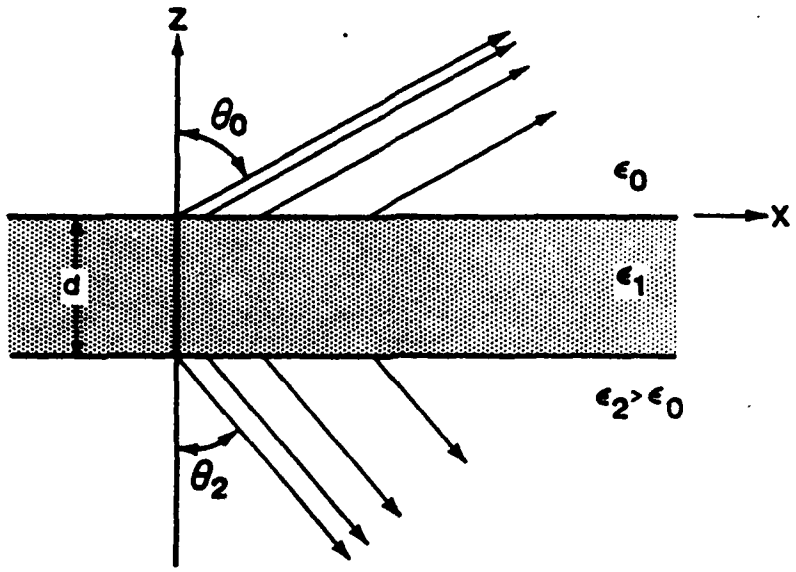
The above discussion has focused on the anomalous-absorption effect because of its peculiar and interesting aspects. However, it must be emphasized that this effect is only one facet of the wave-interaction processes that occur in lossy configurations. As discussed further below, a larger class of wave-interaction phenomena (mostly of the leaky-wave type) represents the broader scope of the present investigation. Because the various wave interactions are well covered by describing the progress made during the past period of the present contract, we list below details of our studies and their more important results.

(a) We have applied the complex pole-zero analysis to examine the reflectance function $\rho(\theta)$ that describes the behavior of a plane-wave incident at an angle θ on a lossy multilayered configuration, as shown in Fig. 1(a). We have found that this analysis distinguishes between two varieties of such configurations, of which one is basically reflecting and the other is basically transmitting. In the absence of absorption losses in any of the media, the former configuration reflects all of the incident plane-wave energy. In contrast, the latter configuration transmits a substantial portion of that energy, such as occurs, for example, if the conducting substrate in Fig. 1(a) is replaced by a dielectric medium. When losses are introduced into any of the media, the analytical and physical behavior of the two configuration varieties are quite different and are best treated separately. In particular, we have shown[26] that the pole-zero analysis, when viewed in the context of a leaky-wave interaction between an incident plane-wave and a layered structure of the basically reflecting type, can produce narrow-band absorption filters designed by using the concept of induced transmission.[12-14]

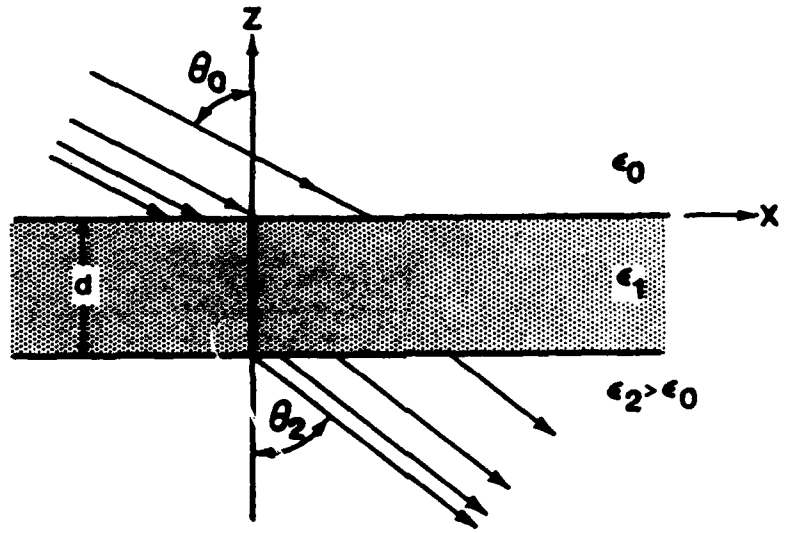
(b) For multilayered configurations of the basically reflecting variety, we have considered their absorption effects in the context of incident Gaussian beams. We have found[26,27] that, when such a beam is incident under critical conditions so as to couple strongly to a leaky wave supportable by the structure, as suggested in Fig. 2, the reflected field exhibits a lateral beam-shift behavior that is anomalous. In that case, the reflected-beam profile can be severely distorted and its shift may appear in a backward rather than the usual forward direction. All of these factors are determined by the amount of absorption losses, as discussed in detail in recent papers.[26-28]

(c) We have performed a preliminary analysis of multilayered configurations of the basically transmitting type so as to obtain their behavior under conditions similar to those discussed in the foregoing item (b). We have thus found that these configurations are of a more complicated nature than those of the basically reflecting type, because the former permit a generally arbitrary partition of the scattered energy into the upper (usually, air) and the lower (substrate) open regions that bound a layered structure. A particularly interesting and novel aspect is that transmitting configurations can support two varieties of leaky waves. A first variety, as shown in Fig. 3(a), is similar to that which occurs in basically reflecting configurations. This first type of leaky waves accounts for radiation into both the upper and lower open regions, and we have shown[29-31] that it can occur in both symmetric (bisectable) and asymmetric configurations. In contrast, as shown in Fig. 3(b), the second variety radiates only in the denser of the two open regions, and these leaky waves can therefore be supported only by asymmetric configurations. The existence of the latter leaky-wave variety has been recognized only recently and, so far, these waves have been explored only for *TE modes*.[24,29-31]

(d) We have extended the analysis of Gaussian-beam incidence described in (b) above to lossless configurations of the basically transmitting type. In this case, the multilayered structures account for a transmitted beam in addition to a reflected one. However, because of the presence of the two possible leaky-wave varieties described in (c) above, the analysis and the phenomenological interaction of the incident beam with the multilayered configuration are different in many respects from those for basically reflecting configurations. We have found that now both the reflected and transmitted beams are displaced and distorted, but their lateral shift and profile distortion are different from those occurring for beams in basically reflecting situations. A particularly interesting aspect is that the transmitted beam always undergoes a forward shift. However, the reflected beam is shifted forward only if incidence occurs from the denser of the two open regions, as shown in Fig. 4(a); in contrast, the reflected beam is shifted backwards if incidence is from the rarer medium, as shown in Fig. 4(b). As the latter is the usual situation for practical cases involving beams incident from air onto a layered structure backed by a dielectric substrate, this result is both novel and intriguing. However, we have so far explored this effect only in lossless configurations,[29-31] and the extension to the lossy case is planned for the next year.

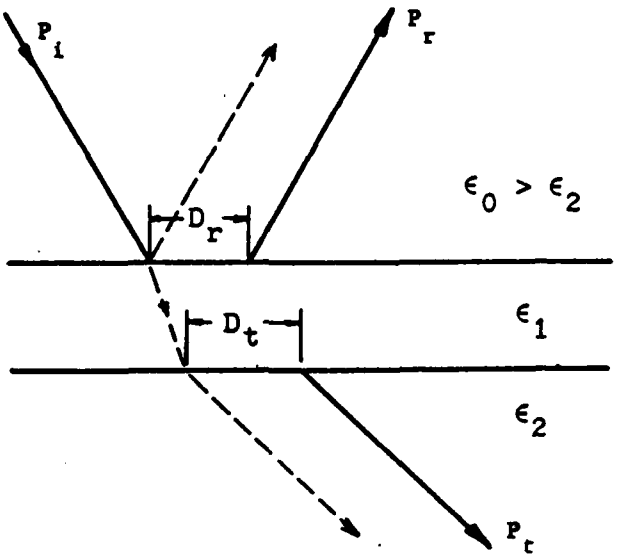


(a) $k_{z2} = +\sqrt{k_{z0}^2 + k^2(\epsilon_2 - \epsilon_0)}$

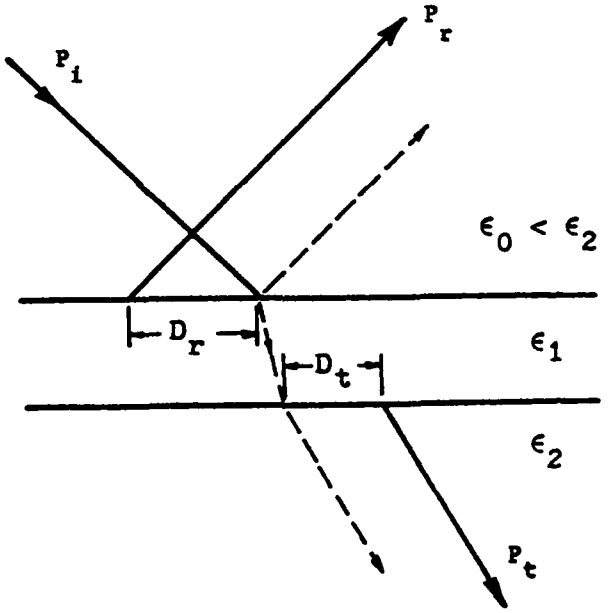


(b) $k_{z2} = -\sqrt{k_{z0}^2 + k^2(\epsilon_2 - \epsilon_0)}$

Fig. 3 Varieties of leaky waves supported by a layer placed between two different dielectric media.



(a) Incidence from denser medium.



(b) Incidence from rarer medium.

Fig. 4 Displacements of reflected and transmitted beams at a basically transmitting configuration. The paths of the reflected and transmitted beam-axes predicted by geometrical optics are shown dashed while the actual ones are shown solid; these define the displacements D_r and D_t, respectively.

5. REFERENCES

1. M.C. Hutley and D. Maystre, "The Total Absorption of Light by a Diffraction Grating," Optics Commun., Vol. 19, pp. 431-436 (December 1976).

2. E.G. Loewen and M. Nevière, "Dielectric Coated Gratings: A Curious Property," Appl. Optics, Vol. 16, pp. 3009-3011 (November 1977).

3. V. Shah and T. Tamir, "Brewster Phenomena in Lossy Structures," Optics Commun., Vol. 23, pp. 113-117 (October 1977).

4. M. Nevière, D. Maystre and P. Vincent, "Determination of the Leaky Modes of a Corrugated Waveguide: Application to the Study of Anomalies," J. Optics (Paris), Vol. 8, pp. 231-242 (1977).

5. V. Shah and T. Tamir, "Simplified Grating Model for the Study of Absorption Anomalies," J. Opt. Soc. Amer., Vol. 69, p. 1473 (October 1979).

6. M. Nevière and P. Vincent, "Brewster Phenomena in a Lossy Waveguide Just Under the Cut-Off Thickness," J. Optics (Paris), Vol. 11, pp. 153-159 (1980).

7. V. Shah and T. Tamir, "Beam Shift Theory of Anomalous Absorption at Leaky-Wave Structures," J. Opt. Soc. Amer., Vol. 70, p. 1606 (December 1980).

8. A. Amittay, P.D. Einziger and T. Tamir, "Experimental Observation of Anomalous Electromagnetic Absorption in Thin-Layered Media," Appl. Phys. Letters, Vol. 38, pp. 754-756 (May 1981).

9. V. Shah and T. Tamir, "Anomalous Absorption by Multi-Layered Media," Optics Commun., Vol. 37, pp. 383-387 (June 1981).

10. L.M. Brekhovskikh, "Waves in Layered Media," (Academic Press, 1980).

11. R. Petit (Editor), "Electromagnetic Theory of Gratings," Topics in Current Physics, Vol. 22 (Springer-Verlag, 1980).

12. P.H. Berning and A.F. Turner, "Induced Transmission in Absorbing Films Applied to Band Pass Filter Design," J. Opt. Soc. Amer., Vol. 47, pp. 230-239 (March 1957).

13. H. Dupoisot, J. Morizet and P. Lostis, "Optical Interference Filters With Large Azimuthal and Spectral Widths," (in French), Appl. Optics, Vol. 13, pp. 1605-1609 (July 1974).

14. P.H. Lissberger, "Coatings with Induced Transmission," Appl. Optics, Vol. 20, pp. 95-104 (January 1981).

15. T. Tamir (Editor), "Integrated Optics," Topics in Applied Physics, Vol. 8 (Springer-Verlag, 1982), Chap. 3, p. 84.

16. K.C. Chang, V. Shah and T. Tamir, "Scattering and Guiding of Waves by Dielectric Gratings with Arbitrary Profiles," J. Opt. Soc. Amer., Vol. 70, pp. 804-813 (July 1980).

17. L.B. Felsen and N. Marcuvitz, "Radiation and Scattering of Waves," (Prentice-Hall, 1973) Chap. 4, p. 370.

18. T. Tamir and H.L. Bertoni, "Lateral Displacement of Optical Beams at Multilayered and Periodic Media," J. Opt. Soc. Amer., Vol. 31, pp. 1397-1413 (October 1971).

19. A. Saad, H. Bertoni and T. Tamir, "Beam Scattering by Non-Uniform Leaky-Wave Structures," Proc. IEEE, Vol. 62, pp. 1552-1561 (November 1974).

20. S.T. Peng, T. Tamir and H.L. Bertoni, "Theory of Periodic Dielectric Waveguides," IEEE Trans. Microwave Theory and Techniques, Vol. MTT-23, pp. 123-133 (January 1975).

21. R.S. Chu and T. Tamir, "Diffraction of Gaussian Beams by Periodically Modulated Media," J. Opt. Soc. Am., Vol. 66, pp. 220-226 (March 1976) and pp. 1438-1440 (December 1976); also, in Vol. 67, pp. 1555-1561 (November 1977).

22. T. Tamir and S.T. Peng, "Analysis and Design of Grating Couplers," Applied Physics, Vol. 14, pp. 235-254 (November 1977).

23. T. Tamir, "Guided-Wave Methods for Optical Configurations," Applied Physics, Vol. 25, pp. 201-210 (July 1981).

24. K. Ogusu, M. Miyagi and S. Nishida, "Leaky TE Modes on an Asymmetric Three-Layered Slab Waveguide," J. Opt. Soc. Amer., Vol. 70, pp. 48-53 (January 1980).

25. V. Shah and T. Tamir, "Leaky-Wave Approach to Induced Transmission", J. Opt. Soc. Amer., Vol. 71, p. 1574 (December 1981).

26. V. Shah and T. Tamir, "Anomalous Absorption Effects for Beams Incident on Lossy Layered Structures", 1982 Joint Intern. IEEE/APS and URSI Symp., Albuquerque, N.M. (May 1982).

27. V. Shah and T. Tamir, "Absorption and Lateral Displacement of Beams Incident on Multilayered Media", J. Opt. Soc. Amer., Vol. 73, pp. 37-44 (January 1983).

28. C.W. Hsue and T. Tamir, "Anomalous Scattering by Layered Media;" National Radio Sci. Meeting, Boulder, Colo. (January 1983).

29. T. Tamir and C.W. Hsue, "Leaky-Wave Interaction Effects for Beams Incident upon Thin-Film Structures", Tech. Dig. 4th Intern. Conf. Integrated Optics and Optical Fiber Commun., pp. 352-353, Tokyo, Japan (June 1983).

30. C.W. Hsue and T. Tamir, "Beam Scattering by a Layered Structure", Proc. 1983 URSI Intern. Symp. Electromagnetic Theory, pp. 191-194, Santiago de Compostela, Spain (August 1983).

31. C.W. Hsue and T. Tamir, "Lateral Displacement of Beams Refracted by Layered Media", to be presented at 1983 meeting of the Opt. Soc. Amer., New Orleans, La. (October 1983).

32. J.M. Zavada, J.J. Fasano, B.H. Li, M.J. Shiau and S.T. Peng, "Strong Absorption of Light by Multilayered Dielectric Metal Structures", Proc. SPIE Meeting, Arlington, VA (April 1983).

6. DoD AND OTHER INTERACTIONS

The studies that form the background of the proposed investigation are mostly of basic theoretical nature. However, because of interactions with other workers in this area at industrial laboratories, our results have stimulated applications-oriented R&D in the recent past. In particular, we have supplied analytical considerations for optical beam couplers and numerical data for infra-red gratings to Dr. P.K. Cheo's group at United Technologies, Hartford, Connecticut. The use of blazing for optimizing the performance of periodic beam couplers, which had been originally proposed by us,[22] has been explored by Dr. Hammer's group at RCA Princeton Laboratories and has stimulated other efforts for designing and fabricating such blazed gratings. Comparisons of analytical procedures and numerical data on beam and periodic-structure problems have occurred continuously with Dr. Streifer of Xerox Corp., Palo Alto, California, as well as with other industrial and academic laboratories. In addition, Professor Tamir has been in touch with Professor Petit's group at the University of Marseille (France) in connection with scattering and guiding by periodic structures, with Professor Gaylord's group at Georgia Institute of Technology in connection with the diffraction of beams by holograms, while consultations and information exchange occur on a frequent basis with other scientists, such as Professor Blok's group at the Delft University of Technology (The Netherlands), Professor Suematsu's group at the Tokyo Institute of Technology (Japan), Professor Ulrich's group at The University of Hamburg-Harburg (W. Germany), and others.

The impact of Professor Tamir's studies can perhaps be best assessed by the large amount of citations to his work in the professional literature. Many of these refer to the book "Integrated Optics" (Springer-Verlag, 2nd printing of 2nd Edition, 1982), which was edited by him and included basic and applied state-of-the-art chapters on that specific area. These and other recent citations have used his results as standards for comparison and refer to his work as a fundamental contribution in the areas of waves along periodic and multilayered media, in opto-acoustic interactions, in scattering of beams by planar interfaces, and others. Most recently, the results described here on the anomalous absorption effect have stimulated applied studies by the Army for utilizing this effect in selective filters for classified purposes.[32]

C. NONUNIFORM OPEN DIELECTRIC WAVEGUIDES: TRANSITIONS AND TAPERS

Professor S.T. Peng

Unit EM3-3

1. OBJECTIVE(S)

A major problem in open dielectric waveguide systems that are under consideration for millimeter-wave integrated circuits is that unwanted scattering of waves can occur at any sharp discontinuity or junction. It is therefore essential that connections between circuit elements in such systems employ gradual transition structures in order to reduce the scattering losses. Some theoretical investigations have been made of such transition structures in open dielectric regions, but they have been quite limited due to the mathematical complexity in analyzing such transition structures in general. We propose to obtain analytical solutions for various types of transitions, tapers and junctions on open dielectric waveguides, by viewing these transition structures as sections of nonuniform waveguides. We intend to employ several different approximate procedures in these analyses, including a new one which was developed by us specifically for this program.

2. APPROACH

The transition structures to be studied will be viewed as nonuniform open dielectric waveguides, and mathematical techniques appropriate to that viewpoint will be employed. In particular, various approximations will be used that are based on the generalized transmission-line equations, cast into traveling-wave form. These various approximations are to be examined critically and their limitations and usefulness assessed, both for known approximate methods and for a new one, which we call the phase-perturbation approach, which is presented here for the first time. The program includes both the study of the validity and utility of the various approximate methods and their application to a number of specific transition structures for open dielectric waveguides.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

We have proved a basic sufficient condition for the convergence of an iteration procedure applied to a system of generalized transmission-line equations that govern the field variations in a nonuniform dielectric waveguide. In order to achieve a systematic and deep understanding of the physical phenomena underlying the scattering of a guided wave by a nonuniform dielectric structure, we have further pursued the solution of the difficult problem from two different directions: one is the continuing application of the method of generalized transmission-line equations to various structures of practical importance, and the other is the search for a mathematical model that renders itself a rigorous analysis.

During the past year, we have successfully formulated, in an exact fashion, the wave propagation in a nonuniform periodic medium. More specifically, it is an almost periodic medium with a two-harmonic modulation, which can be considered as an extension of the well known canonical periodic medium with a sinusoidal modulation. Such a model can be applied to the problem of tapering a periodic dielectric antenna in order to shape the radiating beam and to control the sidelobe level and distribution. Such tapering is usually designed by employing a locally uniform analysis. The approach investigated here should be capable of yielding an exact solution, against which the accuracy of the customary locally-uniform analysis can be compared in specific cases.

This new method, using a combination of a three-term recurrence relation in matrix form and the technique of noncommutative continued fractions, is employed to formulate the problem rigorously. specifically, the new method offers the following advantages: (1) the analysis of the almost periodic medium is in a form analogous to that of the known Mathieu equation for the simpler canonical problem, (2) it applies to both commensurate and noncommensurate cases of the almost periodic medium, and (3) it yields a simple and effective algorithm for accurate numerical computations. Therefore, this method provides a general and unified approach to the solution of wave propagation in an almost periodic medium with a two-harmonic modulation.

Both of these efforts, in two different directions, have yielded important results which were given in two separate presentations to the National Radio Science Meeting,[22,23] at Houston, Texas, in May 1983.

4. STATE OF THE ART AND PROGRESS DETAILS

A. Introduction

An electronic system for communication or radar generally consists of various waveguiding components, electronic devices and antennas, which are often interconnected by gradual transition waveguides, in order to reduce the scattering losses due to junction discontinuities. In open dielectric waveguide systems, such as those currently being studied for potential application to millimeter waves, gradual transition waveguides which serve as interconnections are essential, since scattering of waves can occur at any sharp discontinuity or junction. While a thorough understanding of the characteristics of individual components is the first important step towards the development of a new electronic system, it is, at least, equally important to understand thoroughly the characteristics and performance of an aggregation of interconnected components and devices, so that the system can then be optimally designed. A case in point is the current effort in the development of millimeter-wave systems for communication, guidance and many other purposes; state-of-the-art surveys of such developments were reported in the "Workshop on Millimeter Waves," held in October 1980, at Estes Park, Colorado, and sponsored by the U.S. Army Research Office.

Generally speaking, a transition region between two uniform waveguides, such as a bent or curved waveguide, a taper, or a step-junction discontinuity as a limiting case, may be regarded as a nonuniform waveguide. We propose here to undertake a research program on

nonuniform transition dielectric waveguides, with the structures to be chosen on the basis of potential application to millimeter-wave systems. Analyses of nonuniform dielectric waveguides have been carried out for some specific structures. Circular bends of dielectric slabs and optical fibers have been extensively investigated,[1-6] because of their practical importance and also their special geometry that permits a rigorous analysis in terms of classical functions. Dielectric taper structures have been used and analyzed first as radiating elements for antenna application[7,8] and, more recently, as coupling devices between optical fibers and integrated-optical circuits.[9-13] It is believed that these optical devices will be adapted for use in millimeter-wave integrated circuits and that a larger variety of transition waveguide structures will be needed. Therefore, the previous investigations have been too limited in scope, and a more thorough and systematic understanding of nonuniform open dielectric waveguides is needed so that the circuit components, electronic devices, and antenna structures may be assembled together to perform a specific task.

This research program is intended to supplement the earlier work on uniform open dielectric waveguides for millimeter waves, which has been successfully carried out by us during the past few years and is now in its concluding phase. The uniform dielectric waveguide structures investigated consist of image lines, insular lines, ridge guides, striplines and inverted striplines, etc., which may be given the generic name of "dielectric strip waveguide." Employing the rigorous mode-matching technique, we have analyzed the general guiding characteristics of open dielectric strip waveguides; in particular we have discovered a new and interesting type of leaky wave that can exist on some dielectric strip waveguides under appropriate conditions.[14] Such a new type of leaky wave has been subsequently confirmed by an experiment.[15] A summary of our major research results, including the basic understanding and practical features of uniform dielectric strip waveguides, can be found in a pair of invited companion papers.[16] With the background and mathematical facilities we have built up in the analysis of uniform structures, it is logical and timely for us to move one more step forward to undertake this proposed research program on nonuniform open dielectric waveguides.

Interest in and analysis of nonuniform waveguides (primarily in closed metallic form) have a long history in microwave engineering, and extensive research results have been available in the literature. Complete and thorough reviews on the subject can be found, for example, in the excellent books by Lewin, Chang, and Kuester[5] and Sporleder and Unger.[6] The methods of analysis originally developed for closed nonuniform waveguides may be classified into three main categories:

1. modal approach in terms of generalized transmission-line equations,
2. exact formulation in terms of a curvilinear coordinate system for curved waveguides, and
3. perturbation analyses for structures with gentle non-uniformities.

These methods have all been carried over for the analysis of open dielectric structures,[5,6] but those investigations have been limited in scope.

Among the three methods mentioned above, the first method, that of generalized transmission-line equations, is well known for its general applicability to any nonuniform structure.[16,17-19] Therefore, it is particularly suitable for the analysis of a large class of nonuniform dielectric waveguides from a unified viewpoint. Furthermore, preliminary analysis, to be discussed in detail in the next section, indicates that the method offers the following three advantages:

a) The generalized transmission-line equations are in a form that is convenient for applying the iteration procedure, from which simple analytic, though approximate, solutions can be easily obtained.

b) Each correction term has a familiar physical meaning, so that the wave phenomena taking place in the transition region can be thoroughly explained and interpreted.

c) The accuracy of the results may be systematically improved, if desired.

In the next section, the generalized transmission-line equations are first derived, and then cast into traveling wave form, which is less well known but more useful for both calculational purposes and for physical insight. Several approximate approaches based on these equations are then presented, together with their physical meanings. The amplitude-perturbation approach, both the zero-th order form which leads to the familiar WKB approximation, and the first order form which includes internal reflections due to the nonuniformity, have been known and used previously, particularly the former, of course. The phase-perturbation approach, which is derived and discussed next, is new, and is presented here for the first time. It overcomes an important limitation in the amplitude perturbation approach in that it permits the evaluation of phase changes produced by mode coupling due to nonuniformities instead of having to use only the local phases. Further investigation of the implications and potential of this new procedure will constitute part of the proposed research program.

After the presentation of these approximate methods, a short section is included to indicate why caution must be exercised in connection with the selection of local modes. Finally, the useful staircase-approximation method is described, and it is shown how we have already made use of it to obtain new results in connection with nonuniform dielectric antennas of different taper profiles.[8] That investigation yielded the interesting and originally unexpected result that for optimum performance the taper should end in a sharp point, rather than have a rounded end of some type.

B. Preliminary Analysis of Generalized Transmission-Line Equations for Nonuniform Dielectric Waveguides

Some typical nonuniform dielectric waveguides of practical interest are shown in Figure 1. The structure in Fig. 1(a) shows a taper transition that is intended to optimize the transmission of energy between two uniform waveguides of different sizes. In the limit that the smaller waveguide is absent, as shown in Fig. 1(b), we have a taper that will radiate the guided energy into an open region, and such a

structure is commonly known as a taper radiator. The structure in Fig. 1(c) shows a curved waveguide that connects two nonaligned identical waveguides. More generally, curved waveguides may be assembled together to form a branching circuit, as shown in Fig. 1(d), so that the guided energy may be directed to various parts of a circuit in a desired proportion. Finally, Fig. 1(e) shows a circular bend of a dielectric strip waveguide; such a structure is needed for changing the direction of the guided energy. It is noted that Fig. 1(e) shows the three-dimensional character of the structure, whereas Figs. 1(a)-(d) show only the top view.

In the analysis of uniform open dielectric strip waveguide, there exist two main difficulties that have to be overcome. One is that the proper account for the continuous spectrum in the field representation will lead to a system of coupled integral equations for which no simple solutions can be constructed. The other is that the dielectric strip waveguiding structures are three-dimensional (vector) boundary value problems that require the cross-coupling between constituent TE and TM modes at the dielectric step discontinuities that form the waveguide side walls.[16] Such a vector boundary value problem is rather difficult to handle, even in the special case of uniform dielectric strip waveguides. A physical consequence of such cross-coupling is that some uniform dielectric strip waveguides may become leaky under appropriate conditions.[16] In the case of nonuniform dielectric strip waveguides, the combination of the difficulties associated with the continuous spectrum and the cross-coupling between the TE and TM modes will make the analysis of the class of structures almost intractable. With our extensive experience in dealing with the cross-coupling of modes in uniform dielectric strip waveguides, it will be relatively easy for us to assess the effect of the cross-coupling of modes in nonuniform waveguides and to introduce correction terms, if necessary.

As the first step into this new territory of nonuniform dielectric strip waveguides, that are of particular concern for millimeter-wave integrated circuits, we shall employ the well known effective dielectric constant method,[20,21] which has been shown to be the first-order approximation of an exact mode-matching analysis.[16] Thus, the three-dimensional (vector) boundary-value problem is reduced to a simpler two-dimensional (scalar) boundary-value problem so that the TE and TM modes can be considered separately. Furthermore, since we are concerned with the scattering of guided waves which are confined to the vicinity of the dielectric structure, we shall employ the technique of discretizing the continuous spectrum into a complete set of discrete modes by introducing an oversize metallic waveguide, as is customarily done in the literature. In other words, throughout the present proposed research program, we shall deal explicitly only with two-dimensional structures enclosed in oversize metallic waveguides, so that the mathematical analysis and the physical interpretations of wave phenomena associated with nonuniform dielectric waveguides can be kept simple and clear. It should be stressed that the discretization of modes does introduce some degree of approximation but does not neglect the effect of the continuous spectrum. Therefore, as far as the surface waves are concerned, the presence of an oversize metallic waveguide does not change the physics of the problem. In order to keep the ensuing mathematical analysis tractable, we shall therefore consider in this study only two-dimensional dielectric waveguide structures enclosed by an oversize parallel-plate waveguide.

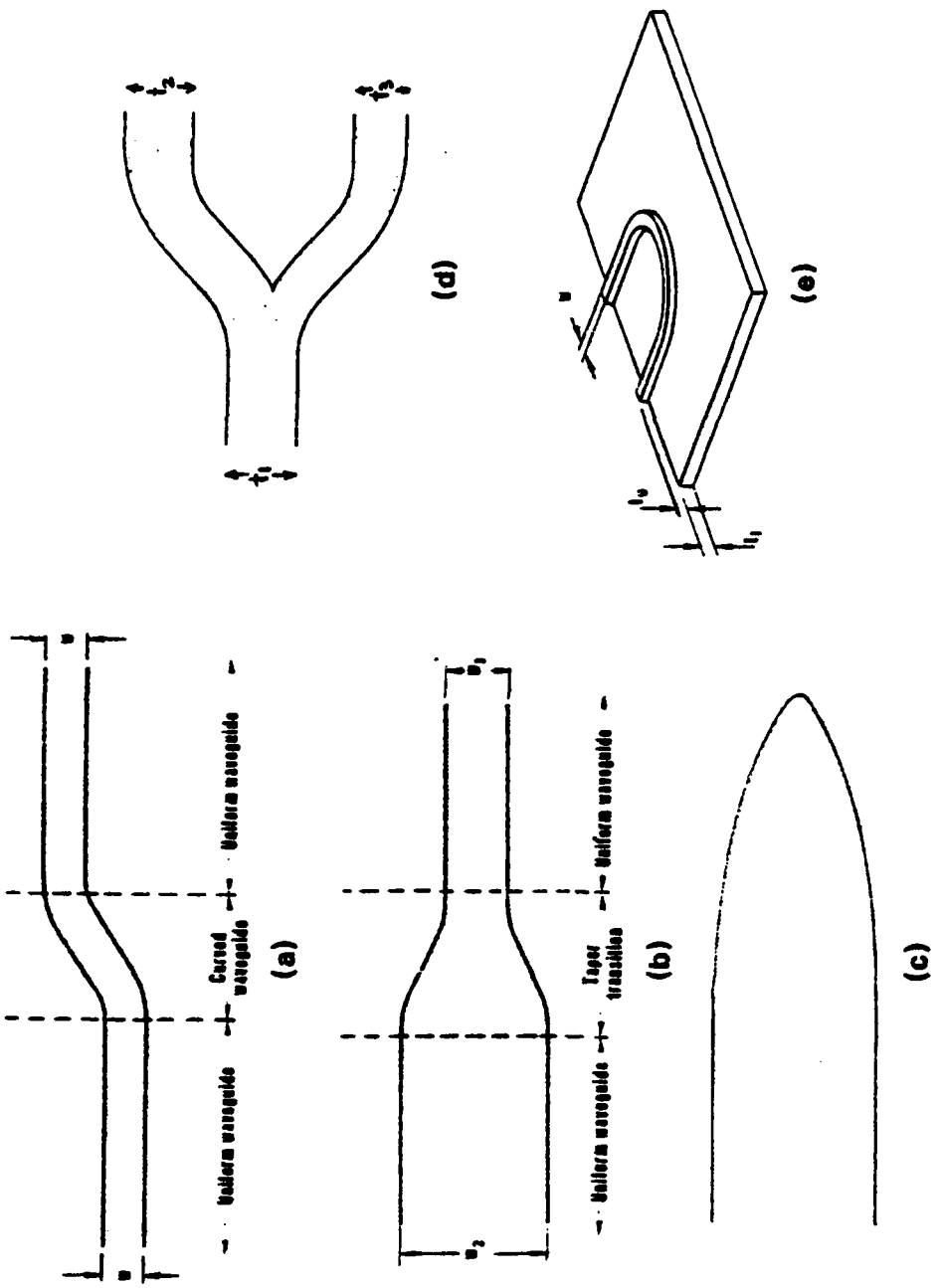


Fig. 1 Typical transition waveguide structures. (a) Curved waveguide between two nonaligned identical waveguides. (b) Taper transition between two unequal waveguides. (c) Taper antenna. (d) Branching circuit. (e) Circular bend.

(1) Derivation of the Generalized Transmission-Line Equations

Here, we review briefly the method of generalized transmission-line equations. A dielectric layer of varying thickness, t(x), is placed on a ground plane, as shown in Figure 2. The structure is invariant along the z direction and we assume that the wave is guided along the x direction. The upper metallic plate at a large height, h, is introduced for the sole purpose of discretizing the waveguide modes; the plate is sufficiently far away that the surface waves are negligibly affected. The complete set of modes of a uniform partially-filled parallel-plate waveguide has been well known; therefore, the complete set of modes at every point x along the dielectric waveguide can be assumed to be well known when we adopt a locally-uniform point of view.

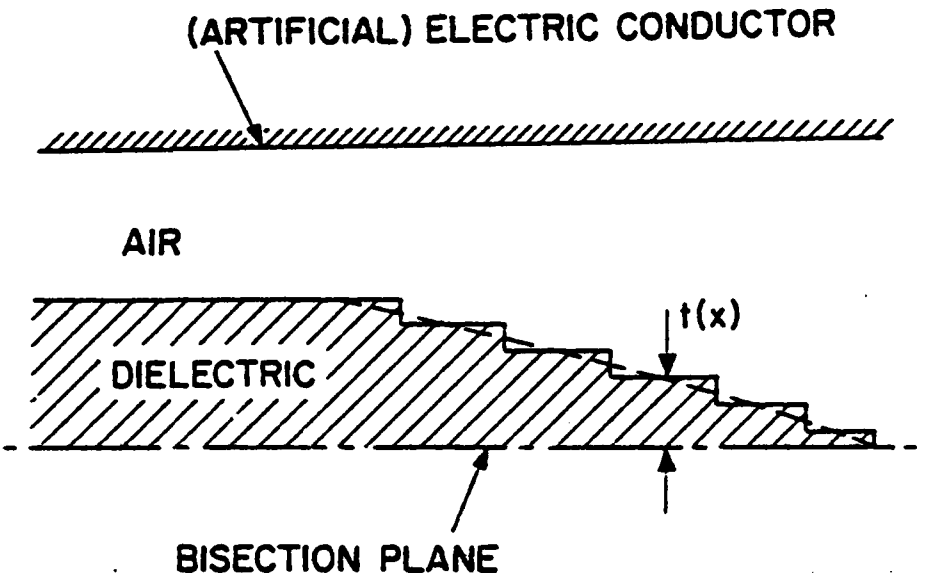


Fig. 2 Nonuniform dielectric waveguide in the form of a taper enclosed by an oversize parallel-plate waveguide as a computational aid. The stepping (staircase approximation) will be explained later, but is not necessary in most of the analyses.

Let the complete set of mode functions at x be denoted by $(\phi_n(x,y), n = 1,2,3,\ldots)$, where $\phi_n(x,y)$ is commonly referred to as the n-th local mode function at x. For a TE wave, for example, the single electric field component may be represented by

$$E_z(x,y) = \sum_{n=1}^{\infty} V_n(x)\, \phi_n(x,y) \qquad (1)$$

where $V_n(x)$ is the modal voltage of the n-th mode. Substituting representation (1) into the Maxwell's equations, we obtain the magnetic field components

$$H_x(x,y) = j \frac{1}{\omega\mu_0} \sum_{n=1}^{\infty} V_n(x) \frac{\partial}{\partial y} \phi_n(x,y) \tag{2}$$

$$H_y(x,y) = - \sum_{n=1}^{\infty} I_n(x) \phi_n(x,y) \tag{3}$$

where $I_n(x)$ is the modal current of the n-th mode, and the generalized transmission-line equations[6,11-13] are

$$\frac{d}{dx} V_m(x) = - j \kappa_m(x) Z_m(x) I_m(x) - \sum_{n=1}^{\infty} F_{mn}(x) V_n(x) \tag{4}$$

$$\frac{d}{dx} I_m(x) = - j \kappa_m(x) Z_m(x) V_m(x) - \sum_{n=1}^{\infty} F_{mn}(x) I_n(x) \tag{5}$$

for $m = 1,2,3,\ldots$. Here, κ_m is the local propagation constant of the m-th mode. Z_m and Y_m are the local wave impedance and admittance, respectively, and they are defined for the TE modes as

$$Z_m(x) = \frac{1}{Y_m(x)} = \frac{\omega\mu_0}{\kappa_m(x)} \tag{6}$$

Finally, F_{mn} is the coefficient of coupling from the n-th to the m-th mode and is defined by the scalar or inner product:

$$F_{mn}(x) = \langle \phi_m(x,y) | \frac{\partial}{\partial x} \phi_n(x,y) \rangle = \int_0^h \phi_m(x,y) \frac{\partial}{\partial x} \phi_n(x,y) dy \tag{7}$$

It can be easily verified from the above definition that F_{mn} possesses the following properties:

$$F_{mn}(x) = 0 \tag{7a}$$

$$F_{mn}(x) = - F_{nm}(x) \tag{7b}$$

for any integers m and n and for any x. Evidently, in the absence of mode coupling ($F_{mn} = 0$), the generalized transmission-line equations (4) and (5) reduce to the ordinary transmission-line equations with nonuniform parameters. It is well known that the nonuniform parameters of a transmission line will cause internal reflections of waves in the longitudinal direction, and that the coupling among modes will result in the redistribution of energy in the transverse plane. The phenomena of internal reflections and redistribution of energy in the transverse plane may degrade the performance of devices in some cases, but may be used to advantages in other cases. Therefore, a thorough and

systematic understanding of these phenomena is essential for the design of nonuniform dielectric waveguides in general and of junction discontinuities in dielectric waveguides in particular. Thus, the key step in the analysis of nonuniform dielectric waveguides is the solution of the generalized transmission-line equations.

Alternatively, we may introduce the transformation of dependent variables, for every m = 1,2,3,...:

$$a_m(x) = \frac{1}{2}\left[Y_m^{\frac{1}{2}}(x)\, V_m(x) + Z_m^{\frac{1}{2}}(x)\, I_m(x)\right] \tag{8}$$

$$b_m(x) = \frac{1}{2}\left[Y_m^{\frac{1}{2}}(x)\, V_m(x) - Z_m^{\frac{1}{2}}(x)\, I_m(x)\right] \tag{9}$$

so that the generalized transmission-line equations, (4) and (5), can be converted into

$$\frac{d}{dx}\, a_m(x) = -\, j\, \kappa_m(x) a_m(x) - \sum_{n=1}^{\infty} S_{mn}(x) a_n(x) - \sum_{n=1}^{\infty} T_{mn}(x) b_n(x) \tag{10}$$

$$\frac{d}{dx}\, b_m(x) = \quad j\, \kappa_m(x) b_m(x) - \sum_{n=1}^{\infty} S_{mn}(x) b_n(x) - \sum_{n=1}^{\infty} T_{mn}(x) a_n(x) \tag{11}$$

for m = 1,2,3,... . Here, the coupling coefficients S_{mn} and T_{mn} are defined by

$$S_{mn}(x) = F_{mn}(x)\left[Y_m^{\frac{1}{2}}(x)\, Z_n^{\frac{1}{2}}(x) + Z_m^{\frac{1}{2}}(x)\, Y_n^{\frac{1}{2}}(x)\right] \tag{12}$$

$$T_{mn}(x) = F_{mn}(x)\left[Y_m^{\frac{1}{2}}(x)\, Z_n^{\frac{1}{2}}(x) + Z_m^{\frac{1}{2}}(x)\, Y_n^{\frac{1}{2}}(x)\right] - \left[\frac{d}{dx} \ln Y_m\right]\delta_{mn} \tag{13}$$

where δ_{mn} is the Kronecker delta. Invoking the property of F_{mn} in (7b), we obtain, from the above definitions:

$$S_{mn}(x) = -\, S_{nm}(x) \tag{12a}$$

$$T_{mn}(x) = T_{nm}(x) \tag{13a}$$

It is well known in the literature that the transformed variables, a_m and b_m, have the physical meanings of the forward and backward traveling waves, respectively. Instead of the original generalized transmission-line equations, (4) and (5), the new equations for the forward and backward traveling waves, (10) and (11), are in a form that is convenient for applying the iteration procedure to obtain approximate analytic solutions for physical interpretations, as further explained below.

(2) Amplitude-Perturbation Approach

From the definitions of the coupling coefficients, S_{mn} and T_{mn}, in (12) and (13), it is easy to verify that these coupling coefficients are small for a structure with weak nonuniformity. Therefore, for a sufficiently smooth structure, the coupling terms can be treated as a small perturbation or neglected altogether. We show here various orders of approximation and indicate their physical significance.

(a) Zero-th-order approximation. In (10) and (11), if we set $S_{mn}(x) = 0$ and $T_{mn}(x) = 0$ for any m and n and for any x, we obtain:

$$\frac{d}{dx}\, a_m^{(0)}(x) = - j\, \kappa_m(x)\, a_m^{(0)}(x) \tag{14}$$

$$\frac{d}{dx}\, b_m^{(0)}(x) = j\, \kappa_m(x)\, b_m^{(0)}(x) \tag{15}$$

where the superscript 0 indicates the zero-th order approximate solutions. The solutions of these two simple equations are:

$$a_m^{(0)}(x) = a_{mo}\, e^{-j\theta_m(x)} \tag{16}$$

$$b_m^{(0)}(x) = b_{mo}\, e^{j\theta_m(x)} \tag{17}$$

$$\text{with } \theta_m(x) = \int_0^x \kappa_m(x)dx \tag{18}$$

where a_{mo} and b_{mo} are arbitrary constants. Evidently, $a_m^{(0)}$ and $b_m^{(0)}$ represent the forward and backward traveling waves. It is noted that $S_{mm} = 0$ but $T_{mm} \neq 0$ in general. Therefore, in setting $S_{mn} = 0$ and $T_{mn} = 0$ for all m and n, we have neglected the couplings among all the modes and also the coupling between the forward and backward traveling modes, so that each mode propagates independently in the nonuniform waveguide without suffering from any internal reflection. Substituting (16) and (17) into (8) and (9) and then adding the resultant equations, we obtain

$$V_m^{(0)}(x) = Y_m^{-\frac{1}{2}}(x)\, [a_{mo}\, e^{-j\theta(x)} + b_{mo}\, e^{j\theta(x)}] \tag{19}$$

for the modal voltage. By invoking (6) and writing explicitly in terms of κ_m, the last equation may be rewritten as:

$$V_m(x) = \frac{\omega\mu_0}{\kappa_m^2(x)} \left[a_{mo}\, e^{-j\int_0^x \kappa_m(\xi)d\xi} + b_{mo}\, e^{j\int_0^x \kappa_m(\xi)d\xi} \right] \qquad (19a)$$

which is recognized as the well known WKB approximate solution. To this order of approximation, the wave propagates according to the local property of the structure. Therefore, the zero-th-order approximation is often referred to as the local-mode approximation.

(b) First-order approximation. If the coupling terms in (10) and (11) are treated as small perturbations, we may apply the iteration procedure to two such coupled equations. For the first-order approximation, we may substitute the unknown solutions, $a_n(x)$ and $b_n(x)$ in the coupling terms by those known solutions of the zero-th-order approximation, $a_n^{(0)}(x)$ and $b_n^{(0)}(x)$, so that the resultant equations can be easily solved. As an illustration, we consider here the special case: $a_{10} \neq 0$; $a_{no} = 0$ for $n \neq 1$; and $b_{no} = 0$ for any n. This is permitted because a_{no} and b_{no} are arbitrary constants and the equations are linear. In doing so, (10) and (11) become:

$$\frac{d}{dx}\, a_m^{(1)}(x) = -\, j\, \kappa_m(x)\, a_m^{(1)}(x) - S_{m1}(x)\, a_{10} e^{-j\theta_1(x)} \qquad (20)$$

$$\frac{d}{dx}\, b_m^{(1)}(x) = \quad j\, \kappa_m(x)\, b_m^{(1)}(x) - T_{m1}(x)\, a_{10} e^{-j\theta_1(x)} \qquad (21)$$

In both of these two equations, the second term is now a known function and can be regarded as an equivalent source term. The particular solutions of these two equations are:

$$a_1^{(1)}(x) = a_{10} e^{-j\theta_1(x)} \qquad (22a)$$

$$a_m^{(1)}(x) = -\, a_{10} e^{-j\theta_m(x)} \int_0^x S_{m1}(\xi)\, e^{j[\theta_m(\xi)-\theta_1(\xi)]}\, d\xi \,, \quad \text{for } m \neq 1 \qquad (22b)$$

$$b_m^{(1)}(x) = -\, a_{10}\, e^{j\theta_m(x)} \int_0^x T_{m1}(\xi)\, e^{-j[\theta_m(\xi)+\theta_1(\xi)]}\, d\xi \,,$$

$$\text{for } m = 1,2,3,\ldots \qquad (22c)$$

In arriving at (22a), we have made use of the fact that $S_{11}(x) = 0$ for any nonuniform structure. The physical interpretation of these solutions is as follows: We start with an incident fundamental forward mode of amplitude a_{10}. Relation (22a) states that the incident fundamental

mode propagates according to the local property of the structure, as in the WKB or local-mode approximation. Relation (22b) states that as the incident fundamental mode propagates along the structure, high forward modes are excited and they propagate according to their own local properties. The excitation coefficient of each mode is given by the integral that depends on the coupling coefficient as well as on the relative phase difference between the mode and the fundamental mode. Relation (22c) states the same effects, except that the modes are propagating in the backward direction. It is worthy noting that unlike $S_{11} = 0$, $T_{11} \neq 0$ and hence $b_1^{(1)}(x) \neq 0$. This means that in the first order approximation, the fundamental backward mode is excited; this accounts for the internal reflections due to the nonuniformity of the structure.

It should be pointed out that for some structures, such as a periodically nonuniform structure, the excitation coefficients as given by the integrals in (22b) and (22c) may become very large or may even blow up under the resonance or phase-matching condition. In that case, the perturbation analysis breaks down and remedial procedures must be introduced. From another viewpoint, the perturbation analysis as described above makes use of the fixed local phase property to determine the amplitudes of the scattered modes, and this is called the amplitude-perturbation approach. When the excitation coefficients are large, the modes are strongly coupled with one another. As a result, the phases of the modes will deviate considerably from those of the local modes and the deviations in the phases can no longer be ignored in an approximate analysis.

(3) Phase-Perturbation Approach

Theoretically, the phase of an independent electromagnetic wave in a multi-mode structure is determined by the eigenvalue of the structure, and the amplitudes of the modes are determined by the eigenvectors. Once an eigenvalue is known, it is then a straightforward matter to determine the corresponding eigenvector. Therefore, knowing the phase information is the key step in the analysis of multi-mode electromagnetic structures, and the amplitude-perturbation approach as described in the preceding section assumes the local phases of the modes as the starting point for the analysis. We outline here an alternative approach, called the phase-perturbation approach, that will permit the evaluation of the phase change due to the mode coupling. The phase-perturbation approach has been known in the literature, but its application to the generalized transmission-line equation is new and appears here for the first time.

For a phase-perturbation analysis, we may apply the Fourier transformation to (10) and (11) and the resultant equations are:

$$-j\kappa A_m(\kappa) = -j K_m(\kappa) * A_m(\kappa) - \sum_{n=1}^{\infty} \sigma_{mn}(\kappa) * A_n(\kappa) - \sum_{n=1}^{\infty} \tau_{mn}(\kappa) * B_n(\kappa) \tag{23}$$

$$-j\kappa B_m(\kappa) = -j K_m(\kappa)*B_m(\kappa) - \sum_{n=1}^{\infty} \sigma_{mn}(\kappa)*B_n(\kappa) - \sum_{n=1}^{\infty} \tau_{mn}(\kappa)*A_n(\kappa) \tag{24}$$

for $m = 1,2,3,\ldots$. Here $A_m(\kappa)$, $B_m(\kappa)$, $K_m(\kappa)$, $\sigma_{mn}(\kappa)$ and $\tau_{mn}(\kappa)$ are the Fourier transforms of $a_m(x)$, $b_m(x)$, $\kappa_m(x)$, $S_{mn}(x)$ and $T_{mn}(x)$, respectively. The symbol * between two functions stands for the convolution of the two functions. To demonstrate the effect of mode couplings on the phase of the fundamental mode, (23) may be written for m = 1 as

$$-j\kappa A_1(\kappa) = -j[K_1(\kappa) + K_1'(\kappa)]*A_1(\kappa) \tag{25}$$

where $K'(\kappa)$ is chosen such that

$$jK_1'(\kappa)*A_1(\kappa) + \sum_{n=1}^{\infty} \sigma_{mn}(\kappa)*A_n(\kappa) + \sum_{n=1}^{\infty} \tau_{mn}(\kappa)*B_n(\kappa) = 0 \tag{26}$$

Taking the inverse Fourier transformation of the last equation, we obtain

$$j\kappa_1'(x)a_1(x) + \sum_{n=1}^{\infty} S_{mn}(x)a_n(x) + \sum_{n=1}^{\infty} T_{mn}(x)b_n(x) = 0 \tag{27}$$

Utilizing the amplitude-perturbation results, (22), as an approximation for the a's and b's, we obtain

$$j\kappa_1'(x) = \sum_{n=1}^{\infty} S_{mn}(x)\, e^{-j[\theta_m(x)-\theta_1(x)]} \int_0^x S_{mn}(\xi)\, e^{j[\theta_m(\xi)-\theta_1(\xi)]}\, d\xi$$

$$+ \sum_{n=1}^{\infty} T_{mn}(x)\, e^{j[\theta_m(x)+\theta_1(x)]} \int_0^x T_{mn}(x)\, e^{-j[\theta_m(\xi)+\theta_1(\xi)]}\, d\xi \tag{28}$$

which is a known function for a given nonuniform structure. The solution of (25) can now be written as:

$$a_1(x) = a_{10}\, e^{-j\int_0^x [\kappa_1(\xi)-\kappa_1'(\xi)]d\xi} \tag{29}$$

Evidently, κ_1' represents the effect of the higher modes on the local phase constant of the structure.

The key to the phase-perturbation analysis is the solution of (23) and (24). The approximate solution given above requires further examination, and possibly a better solution remains to be explored. Part of the proposed research program is addressed to this question; in general, we intend to investigate the implications and potential of this new approach.

(4) Concept of Local Mode Functions

By way of a simple example for which the exact solution is known, we show here that great care should be exercised in determining a complete set of local mode functions. Otherwise, one may be led to erroneous results. The simple example is the scattering of a guided wave by a corner of a parallel-plate metallic waveguide, as shown in Figure 3. By symmetry, the boundary condition on the plane containing the corner will be that of a perfect magnetic conductor or a perfect electric conductor, for symmetric or antisymmetric excitation, respectively. Therefore, the only problem here is to determine a general solution for the two arms of the structure, each being a parallel-plate waveguide inclined at an angle with respect to the coordinate system of the structure.

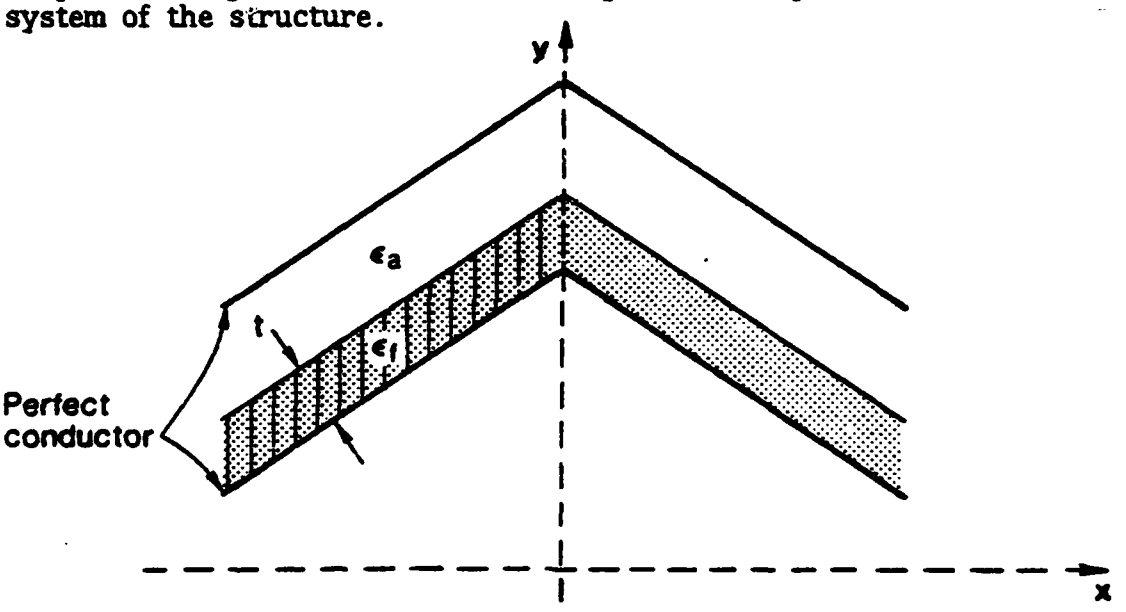


Fig. 3 Corner junction between two uniform waveguides.

A parallel-plate waveguide of a given height $\bar{h}$ is inclined at an angle θ with respect to the xz-plane, as shown in Figure 4. At any x, the waveguide has a height h related to $\bar{h}$ by

$$h = \bar{h}/\cos\theta \tag{30}$$

which is independent of x in this particular case. The coordinates of the two parallel plates are given by

$$y - x\tan\theta = 0 \tag{31a}$$

$$y - x\tan\theta = h \quad . \tag{31b}$$

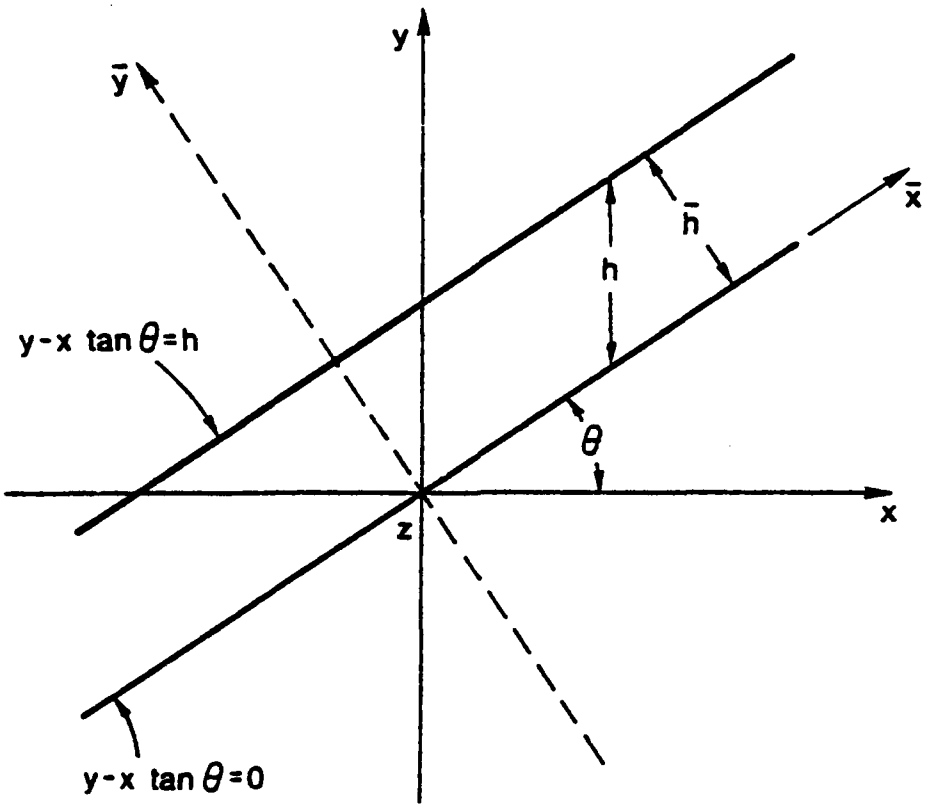


Fig. 4 Parallel-plate waveguide in the structure coordinate system.

The local mode functions are those of a parallel-plate waveguide of the height h; they are given by

$$\phi_n(x,y) = \sqrt{\frac{2}{h}} \sin \frac{n\pi}{h}(y-x\tan\theta) \ , \qquad n = 1,2,3,... \tag{32}$$

for the TE modes and at any x. With such a set of mode functions, we have the coupling coefficients

$$P_{mn} = \langle \phi_m(x,y) | \frac{\partial}{\partial x} \phi_n(x,y) \rangle$$

$$= -\frac{2}{h}\frac{n\pi}{h}\tan\theta \int_{x\tan\theta}^{h+x\tan\theta} \sin\frac{m\pi}{h}(y-x\tan\theta)\cos\frac{n\pi}{h}(y-x\tan\theta)dy$$

$$= 0 \tag{33}$$

for any m and n and for any x. This means that no coupling occurs among the local modes and such a result is known to be incorrect for the simple case under consideration, as explained below.

It is well known that for a TE wave, the general solution for the electric field in the parallel-plate waveguide can be represented in terms of the mode functions in the natural coordinate system $\bar{x}$, $\bar{y}$ and z of the waveguide as:

$$E_z(\bar{x},\bar{y}) = \sqrt{\frac{2}{h}} \sum_{n=1}^{\infty} [\bar{a}_n e^{-j\bar{\kappa}_n\bar{x}} + \bar{b}_n e^{j\bar{\kappa}_n\bar{x}}] \sin\frac{n\pi}{h}\bar{y} \tag{34}$$

By invoking the formula for coordinate rotation:

$$\bar{x} = x\cos\theta + y\sin\theta , \tag{35}$$

$$\bar{y} = y\cos\theta - x\sin\theta , \tag{36}$$

the general solution may be expressed in terms of the structure coordinate system as

$$E_z(x,y) = \sum_{n=1}^{\infty} [\bar{a}_n e^{-j\bar{\kappa}_n x\cos\theta} e^{-j\bar{\kappa}_n y\sin\theta} + \bar{b}_n e^{j\bar{\kappa}_n x\cos\theta} e^{j\bar{\kappa}_n y\sin\theta}] \sin\frac{n\pi}{h}(y-x\tan\theta)$$

$$= \sum_{n=1}^{\infty} [a_n e^{-j\kappa_n x} \phi_n^+(x,y) + b_n e^{j\kappa_n x} \phi_n^-(x,y)] \tag{37}$$

where ϕ_n^+ and ϕ_n^- can be interpreted as the n-th local mode functions for the forward and backward traveling waves, respectively, and are defined by:

$$\phi_n^{\pm}(x,y) = \sqrt{\frac{2}{h}} e^{\mp j\kappa_n y} \sin\frac{n\pi}{h}(y-x\tan\theta) . \tag{38}$$

A comparison of $\phi_n^{\pm}(x,y)$ above with $\phi_n(x,y)$ in (32) shows that each local mode function must include a y-dependent phase factor that was not anticipated. This shows that, in some cases, the concept of local modes has to be modified or great care has to be exercised in defining the local mode functions.

(5) The Method of Staircase Approximation

A nonuniform dielectric waveguide of continuous geometrical profile is shown in Figure 5. The scattering of surface waves by that type of structure is not amenable to an exact analysis even for simple geometrical profiles and one must resort to approximate analysis. A commonly-employed approximation technique is to replace the continuous profile by a piecewise-constant one, as shown by the dashed line in Fig. 5, and this is known as the staircase approximation. Evidently, in the limit

when the step size is reduced to zero, the piecewise-constant profile will approach the continuous one. It can be shown that such a limiting process can be taken to derive the generalized transmission-line equations for the analysis of the nonuniform dielectric waveguides.

With the piecewise-constant profile, the structure consists of a cascade of uniform constituent regions separated by step discontinuities. Each constituent region can be considered as a portion of a uniform partially-filled parallel-plate waveguide for which a complete set of discrete eigenvalues and their corresponding mode functions are well known. The electromagnetic fields in each constituent region may then be represented in terms of the complete set of mode functions of the region, and they are required to satisfy the boundary conditions at each step discontinuity. To simplify the formulation, the staircase structure may be viewed as a cascade of basic units, each consisting of a step discontinuity and a uniform waveguide of finite length, and the scattering of surface waves by the staircase structure can then be analyzed in terms of the scattering by each basic unit. Since the scattering of surface waves by an individual basic unit is much simpler to handle, it can be considered as a constituent problem or a building block for the analysis of the whole staircase structure with relative ease.

The scattering of surface waves by a step discontinuity between two uniform dielectric waveguides had been rigorously formulated and analyzed by the author as a constituent problem for the study of dielectric strip waveguides.[16] It was shown that the coupling of modes at a step discontinuity can be represented by an ideal transformer for

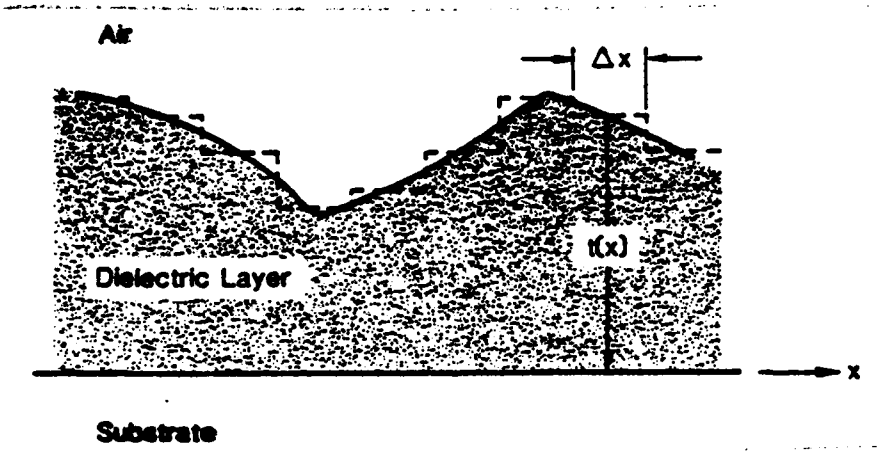


Fig. 5 Staircase approximation for a nonuniform dielectric waveguide of continuous profile.

which the transfer matrix is well established. In the finite uniform waveguide region, each mode propagates independently and may be represented by a transmission-line section for which the transfer matrix is well known. Thus, the transfer matrix for a basic unit is simply determined by the product of the transfer matrices for the step discontinuity and the uniform waveguide section. Moreover, the overall transfer matrix of the staircase structure can in turn be obtained as the simple product of those of the basic units; the scattering of surface waves by the staircase structure can therefore be considered to have been solved.

From the outline above, the method of staircase approximation is seen to be very simple and straightforward. It applies to structures with any geometrical profile and any material composition. In addition, the building-block approach with the addition of the transfer-matrix technique is particularly useful for the development of computer programs for numerical analysis. Therefore, this method is particularly suitable for a comparative study of a large class of nonuniform dielectric waveguides. In fact, this method had been employed by us for the study of the performance of dielectric taper antennas.[8] A computer program for the scattering of surface waves by dielectric taper structures had been successfully developed and many interesting results had been obtained and reported.[8] For example, Fig. 6 shows the effect of taper profile on the reflection of surface waves from the antenna structure. Evidently, the reflected power can be quite substantial, if the taper antenna is not properly designed. To our knowledge, such a reflection phenomenon had been totally neglected in the previous analyses of dielectric taper antennas; the staircase approximation permits the evaluation of its effect for the first time. These curves for three different taper profiles permit us to suggest that while the aspect ratio of a taper should be as large as possible, an optimum profile should have a smooth entrance end, as in the case of the elliptic taper, and a needle-sharp exit end, as in the case of a linear taper. Such an interesting result is unexpected in the light of previous analyses that did not take into account the internal reflection of waves by the continuously varying tapers. Since the staircase profile is an approximation of a continuous one, the numerical results we have already obtained for the dielectric taper antennas can be used, at least, as a guideline for the construction of analytical formulas by the method of the generalized transmission-line equations in the presently proposed research program.

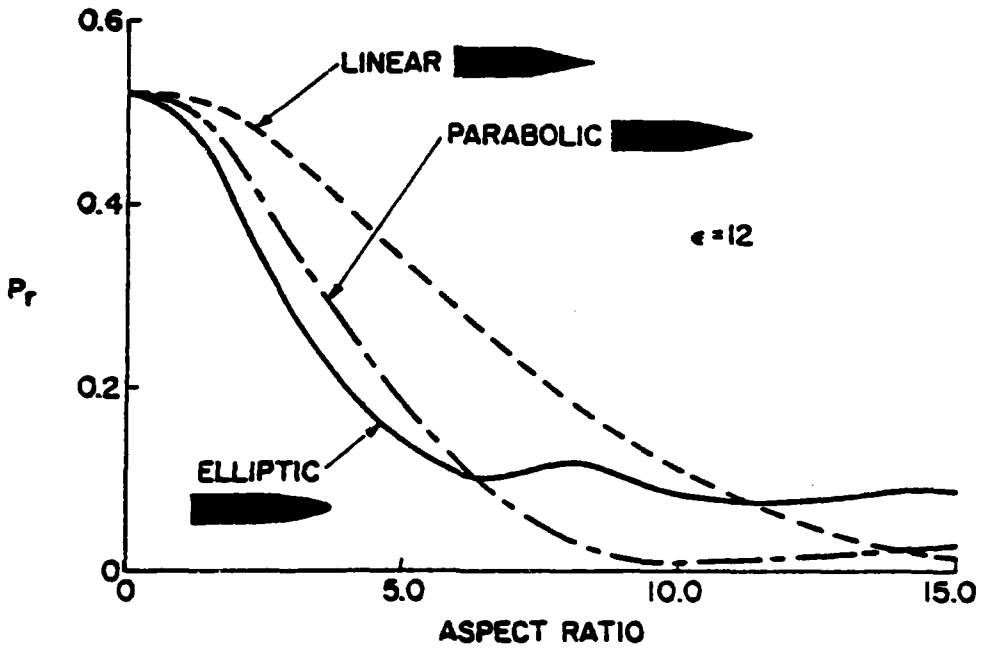


Fig. 6 Power reflected from different dielectric taper antennas as a function of aspect ratio.

C. Convergence of the Iteration Procedure for Generalized Transmission-Line Equations

We have already made some significant progress in laying down a solid mathematical foundation for the analysis of nonuniform transition waveguides by the method of generalized transmission-line equations. It is recalled that the governing equations for the mode amplitudes, (10) and (11), are infinite systems of first order differential equations with variable coefficients. For such systems of differential equations, in general, no exact solution can be expected, and one should be very satisfied with an approximate solution with known accuracy. In practice, an infinite system of differential equations has to be truncated to a finite system for an approximate analysis. In the present case of differential equations with variable coefficients, however, the truncation alone will not permit an accurate analysis afterwards, unless it is truncated to a single mode approximation or unless the effect of mode coupling is totally neglected. We have observed that, by converting the differential equations into integral equations, various orders of approximate solutions can be obtained by the iteration procedure. We have established a sufficient condition for the convergence of the iteration process, as summarized below.

Equations (10) and (11) are two systems of first order differential equations which can be converted into systems of integral equations. The systems of integral equations may take many different forms, each of which may be more suitable for some purposes than the others. After a careful study, we have determined that for the proof of convergence, it is necessary to cast the integral equations in the following form:

$$\underline{a}(x) = e^{-j\theta(x)} \left\{ \underline{a}(o) - \int_o^x e^{j\theta(\xi)} \left[S(\xi)\underline{a}(\xi) + T(\xi)\underline{b}(\xi) \right] d\xi \right\} \tag{39}$$

$$\underline{b}(x) = e^{+j\theta(x)} \left\{ e^{-j\theta(L)}\, \underline{b}(L) - \int_x^L e^{-j\theta(\xi)} \left[S(\xi)\underline{a}(\xi) + T(\xi)\underline{b}(\xi) \right] d\xi \right\} \tag{40}$$

where $\underline{a}$ and $\underline{b}$ are column vectors, S and T are known matrices with the general elements S_{mn} and T_{mn} defined by (12) and (13), and θ is a diagonal matrix with θ_m, defined by (18), at the m-th diagonal position. It is important to note that the ranges of integration $\underline{a}$ and $\underline{b}$ in the last two equations are complementary to each other within the nonuniform region, $0 \leq x \leq L$. Physically, $\underline{a}(x)$ represents the forward traveling waves, and the contribution to $\underline{a}(x)$ must be from the entrance end to the field point x. On the other hand, $\underline{b}(x)$ represents the backward traveling waves due to the reflections from the exit end and the contribution to $\underline{b}(x)$ must be from the exit end to the field point x. Therefore, the basic physical processes of the interaction of the electromagnetic waves with the nonuniformity of the structure have been included in the basic equations from which the field quantities are determined. It turns out that the particular form of the integral equations, (39) and (40), is not only physically plausible, but also mathematically necessary for the convergence of the solutions, as explained next.

By the iteration procedure, we may start the zero-th order approximate solutions of (30) and (40):

$\underline{a}(x)$ represents the forward traveling waves, and the contribution to $\underline{a}(x)$ must be from the entrance end to the field point x. On the other hand, $\underline{b}(x)$ represents the backward traveling waves due to the reflections from the exist end and the contribution to $\underline{b}(x)$ must be from the exist end to the field point x. Therefore, the basic physical processes of the interaction of the electromagnetic waves with the nonuniformity of the structure have been included in the basic equations from which the field quantities are determined. It turns out that the particular frm of the integral equations, (39) and (40), is no only physical plassible, but also mathematically necessary for the convergence of the solutions, as explained next.

By the iteration procedure, we may start the zero-th order approximate solutions of (39) and (40):

$$\underline{a}^{(o)}(x) = e^{-j\theta(x)}\ \underline{a}(o) \tag{41}$$

$$\underline{b}^{(o)}(x) = e^{-j[\theta(L)-\theta(x)]}\ \underline{b}(L) \tag{42}$$

and then generate the n-th order approximate solutions:

$$\underline{a}^{(n)}(x) = e^{-j\theta(x)}\left\{\underline{a}(o) - \int_o^x e^{j\theta(\xi)}\left[S(\xi)\underline{a}^{(n-1)}(\xi) + T(\xi)\underline{b}^{(n-1)}(\xi)\right]d\xi\right\} \tag{43}$$

$$\underline{b}^{(n)}(x) = e^{-j\theta(x)} = e^{-j\theta(L)}\left\{\underline{b}(L) - \int_x^L e^{j\theta(\xi)}\left[S(\xi)\underline{b}^{(n-1)}(\xi) + T(\xi)\underline{a}^{(n-1)}(\xi)\right]d\xi\right\} \tag{44}$$

Taking the difference of two successive approximations, we obtain from the last two equations:

$$\underline{a}^{(n)}(x)-\underline{a}^{(n+1)}(x) = -e^{-j\theta(x)}\int_o^x e^{j\theta(\xi)}\left\{S(\xi)\left[\underline{a}^{(n-1)}(\xi)-\underline{a}^{(n)}(\xi)\right] + T(\xi)\left[\underline{b}^{(n-1)}(\xi)-\underline{b}^{(n)}(\xi)\right]\right\}d\xi \tag{45}$$

$$\underline{b}^{(n)}(x)-\underline{b}^{(n+1)}(x) = -e^{j\theta(x)} \int_x^L e^{-j\theta(\xi)}\left\{S(\xi)\left[\underline{b}^{(n-1)}(\xi)-\underline{b}^{(n)}(\xi)\right]+ T(\xi)\left[\underline{a}^{(n-1)}(\xi)-\underline{a}^{(n)}(\xi)\right]\right\}d\xi \quad (46)$$

By the Schwartz inequality, we then obtain, in the norms of the sectors and matrices:

$$||\underline{a}^{(n)}(x)-\underline{a}^{(n+1)}(x)|| \leq ||S(x)||\ ||\underline{a}^{(n-1)}(x)-\underline{a}^{(n)}(x)||+ ||T(x)||\ ||\underline{b}^{(n-1)}(x)-\underline{b}^{(n)}(x)|| \quad (47)$$

$$||\underline{b}^{(n)}(x)-\underline{b}^{(n+1)}(x)|| \leq ||S(x)||\ ||\underline{b}^{(n-1)}(x)-\underline{b}^{(n)}(x)||+ ||T(x)||\ ||\underline{a}^{(n-1)}(x)-\underline{a}^{(n)}(x)|| \quad (48)$$

Based on these inequalities, the iteration procedure as described by (41) - (44) converges, if the matrices S and T satisfy the following condition:

$$||S(x)|| + ||T(x)|| = \rho<1 \quad (49)$$

Thus, we have established a sufficient condition for the convergence of the iterative solutions, (41) - (44). For a given nonuniform structure, S and T can be readily determined and we are now in the process of examining this sufficient condition for various nonuniform waveguides of practical interest. As an example, in the limit of uniform waveguide, we have: S=o and T=o. In such an extreme case, we obtain from (49): ρ=o, and the zero-th order approximate solutions (41) and (42) are, in fact, the exact solutions, as should be expected.

The proof of the sufficient condition for the convergence of the iteration procedure, and also the results of specific examples, have been presented at the National Radio Science Meeting[22] in May 1983.

D. Wave Propagation in an Almost Periodic Medium

Consider a nonuniform medium characterized by the dielectric constant:

$$\varepsilon(x) = \varepsilon_{ave}\ (1 + 4\delta \cos\frac{2\pi}{a}\ x \cos\frac{2\pi}{b}\ x) \quad (50)$$

where ε_{ave} is the average dielectric constant, δ is the modulation index, and a and b are the two periods. Without loss of generality, we assume b > a throughout this work. Such a medium can be regarded as a periodic medium with a periodic modulation index; thus, a is the basic period and b is the modulation period that characterizes the non-

uniformity of the periodic medium. If a and b are commensurate, the medium again becomes periodic with their largest common factor as the period; otherwise, it is nonperiodic. However, when b >> a, the medium is almost periodic, meaning that we have a slightly nonuniform periodic medium that corresponds to slowly tapered periodic antenna, and it is such a nonuniformity that can be utilized for suitably tapering the antenna aperature distribution to produce the desired radiation pattern.

The analysis of wave propagation in a nonuniform medium is a difficult problem, but we observe that the medium with two periods can be rigorously treated by the technique of generalized continued fractions. Therefore, we have a mathematical model for which reliable results with any degree of desired accuracy can be obtained, and such results can be used as a standard against which other approximate ones can be compared and judged.

The propagation of waves in the dielectric medium characterized by (50) is governed, for TE modes, by:

$$\left[\frac{d^2}{dx^2} + (k^2 - k_t^2) + 4\delta k^2 \cos \frac{2\pi}{a}x \cos \frac{2\pi}{b}x\right] E(x) = 0 \tag{51}$$

where $k = k_o\sqrt{\varepsilon_{ave}}$ is the average wavenumber of the plane wave in the medium, k_t is the transverse wavenumber, and $E(x)$ characterizes the longitudinal variation of the electric field. We may represent the field solution by the double Fourier series:

$$E(x) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} E_{mn}\, e^{-jk_{xmn}x} \tag{52}$$

with

$$k_{xmn} = k_x + \frac{2m\pi}{a} + \frac{2n\pi}{b} \tag{53}$$

where k_x is the characteristic wavenumber and E_{mn} is the mn-th space harmonic amplitude. Substituting the field representation into the governing equation, we obtain the recurrence relation:

$$(k^2-k_t^2-k_{xmn}^2)E_{mn} + \delta k^2(E_{m+1,n-1} + E_{m-1,n+1} + E_{m+1,n+1} + E_{m-1,n-1}) = 0 \tag{54}$$

which is a second-order partial difference equation with variable coefficients. Effecting the change of variable, $m = \ell+n$, and setting $V_{\ell,n} = E_{\ell+n,n}$, (54) may be written in matrix form as:

$$\delta J\underline{V}_{\ell+2} + D_\ell\underline{V}_\ell + \delta J^T\underline{V}_{\ell-2} = 0 \tag{55}$$

where J is a matrix with 1 at every entry of the first right off-diagonal and 0 elsewhere; J^T is the transpose of J; and D_ℓ is a matrix with

$(1-(k_t^2+k^2_{x(\ell+n)n})/k^2)$ at the n-th diagonal entry, 1 at the first off-diagonal entries and 0 elsewhere. Equation (55) may also be viewed as a second order vector difference equation for which there exist two independent solutions:

$$V_\ell^{(1)} = K_\ell^+ \; V_{\ell-2}^{(1)} \tag{56}$$

$$V_\ell^{(2)} = K_\ell^- \; V_{\ell+2}^{(2)} \tag{57}$$

where K_ℓ^+ and K_ℓ^- are the continued fractions of ascending and descending order, respectively, and are given by:

$$K_\ell^+ = \delta(D_\ell - \delta^2 J(D_{\ell+2} - \delta^2 J(D_{\ell+4} - \ldots)^{-1} J^T)^{-1} J^T)^{-1} J^T \tag{58}$$

$$K_\ell^- = \delta(D_\ell - \delta^2 J^T(D_{\ell-2} - \delta^2 J^T(D_{\ell-4} - \ldots)^{-1} J)^{-1} J)^{-1} J \tag{59}$$

A sufficient condition for the convergence of the continued fractions with non-commutative elements, such as the matrices in this case, has been established previously by the author. In view of the factor $\delta^2 < 1$ in the partial numerators, the continued fractions converge rapidly, providing for an effective technique for both analytic and numerical investigation. A computer program based on the continued-fraction algorithm has been developed and interesting results have been obtained during the past year. Some of the results were presented at the National Radio Science Meeting[23] in May 1983.

5. REFERENCES

1. E.A.J. Marcatili, "Bends in Optical Dielectric Guides," Bell Syst. Tech. J., Vol. 48, pp. 2103-32 (1969).

2. D. Marcuse, "Bending Losses of the Asymmetric Slab Waveguide," Bell Syst. Tech. J., Vol. 50, pp. 2551-63 (1971).

3. L. Lewin, "Radiation from Curved Dielectric Slabs and Fibers," IEEE Trans. MTT, Vol. MTT-22, pp. 718-727 (1974).

4. D.C. Chang and E.F. Kuester, "Radiation and Propagation of a Surface-Wave Mode on a Curved Open Waveguide of Arbitrary Cross Section," Radio Science, Vol. 11, pp. 449-457 (1976).

5. L. Lewin, D.C. Chang, and E.F. Kuester, Electromagnetic Waves and Curved Structures, Peter Peregrinus Ltd., Stevenage, England (1977).

6. F. Sporleder and H.G. Unger, Waveguide Tapers, Transitions and Couplers, Peter Peregrinus Ltd., Stevenage, England (1979).

7. D.G. Keily, Dielectric Aerials, John Wiley and Sons, New York (1953).

8. S.T. Peng and F. Schwering, "Effect of Taper Profile on Performance of Dielectric Taper Antennas," Proc. National Radio Science Meeting, University of Washington, Seattle, WA, p. 96 (June 1979).

9. R.K. Winn and J.H. Harris, "Coupling from Multimode to Single-Mode Linear Waveguides Using Horn-Shaped Structures," IEEE Trans. MTT, Vol. MTT-23, pp. 92-97 (1975).

10. A.R. Nelson, "Coupling Optical Waveguides by Tapers," Appl. Opt., Vol. 14, pp. 3012-3015 (1975).

11. W.K. Burns, A.F. Milton and A.B. Lee, "Optical Waveguide Parabolic Coupling Horns," Appl. Phys. Lett., Vol. 30, pp. 28-30 (1977).

12. J.C. Campbell, "Tapered Waveguide for Guided Wave Optics," Appl. Opt., Vol. 18, pp. 900-902 (1979).

13. D. Marcuse, "Radiation Losses of Step-Tapered Channel Waveguides," Appl. Opt., Vol. 19, pp. 3676-2681 (1980).

14. S.T. Peng and A.A. Oliner, "Leakage and Resonance Effects on Strip Waveguides for Integrated Optics," Trans. of Electronics and Communication Engineers of Japan, Special Issue on Integrated Optics and Optical Fiber Communications, Vol. E61, pp. 151-154 (March 1978).

15. K. Ogusu and I. Tanaka, "Optical Strip Waveguide: An Experiment," Appl. Opt., Vol. 19, pp. 3322-3325 (1 October 1980).

16. (a) S.T. Peng and A.A. Oliner, "Guidance and Leakage Properties of a Class of Open Dielectric Waveguides, Part I: Mathematical Formulation," IEEE Trans. MTT, Special Issue on Open Dielectric Waveguides (September 1981).

(b) A.A. Oliner, S.T. Peng, T.I. Hsu and A. Sanchez, "Guidance and Leakage Properties of a Class of Open Dielectric Waveguides, Part II: New Physical Effects," IEEE Trans. MTT, Special Issue on Open Dielectric Waveguides (September 1981).

17. E. Bahar, "Radio Wave Propagation in Stratified Media with Nonuniform Boundaries and Varying Electromagnetic Parameters: Full Wave Analysis," Canadian J. Phys., Vol. 50, pp. 3132-42 (1972).

18. E. Bahar, "Coupling Between Guided Surface Waves, Lateral Waves, and the Radiation Fields by Rough Surfaces -- Full-Wave Solutions," IEEE Trans. MTT, Vol. MTT-25, pp. 923-931 (1977).

19. E. Bahar, "Excitation of Surface Waves and the Scattered Radiation Fields by Rough Surfaces of Arbitrary Slope," IEEE Trans. MTT, Vol. MTT-28, pp. 999-1006 (1980).

20. R.M. Knox and P.P. Toulios, "Integrated Circuits for the Millimeter Wave Through Optical Frequency Range," Proc. Symp. on Submillimeter Waves, Polytechnic Press of Polytechnic Institute of Brooklyn, pp. 497-516 (1970).

21. W.V. McLevige, T. Itoh and R. Mittra, "New Waveguide Structures for Millimeter-Wave and Optical Integrated Circuits," IEEE Trans. Microwave Theory Tech., Vol. MTT-23, pp. 788-794 (1975).

22. S.T. Peng and F. Schwering, "Method of Analysis for Nonuniform Dielectric Waveguides," Digest of National Radio Science Meeting, p. 104, Houston, TX (May 23-26, 1983).

23. A. Cheo and S.T. Peng, "Rigorous Analysis of Wave Propagation in an Almost Periodic Medium," Digest of National Radio Science Meeting, p. 33, Houston, TX (May 23-26, 1983).

6. DoD AND OTHER INTERACTIONS

Professor Peng spent two summers (1977, 1978) at the US Army CORADCOM at Fort Monmouth, New Jersey, working with Dr. Felix Schwering on dielectric grating and taper antennas, under an LRCP arrangement with the US Army Research Office. These interactions have continued in the form of a Post-LRCP contract and successor contracts.

D. COLLECTIVE FORMULATION OF WAVE PHENOMENA FOR GUIDING AND TRANSMISSION

1. OBJECTIVE(S)

High frequency propagation in, or transmission through, layered media usually requires synthesis in terms of a large number of basic wave processes. For the guiding or ducting problem, these wave processes are either normal (discrete and continuous) modes or ray-optical fields. For the transmission problem, the basic wave processes are traveling waves which undergo multiple internal reflection at the layer boundaries. Because descriptions by multiple propagation events are usually poorly convergent and do not provide physical interpretation in compact form, it is desirable to seek collective descriptions of multiple phenomena.

We shall explore the rigorous formulation of a group of propagation events of one type in terms of a more compact collective representation of the same or of another type. By dealing with a group of events extracted from the totality, one may develop hybrid representations that retain the original events where these are convenient and meaningful but employ collective alternatives of another type when these are more appropriate. Moreover, one may seek a formulation wherein fundamental multiple propagation events in a complicated medium are re-expressed collectively as events of the same type in an equivalent composite medium.

The preceding remarks refer to fields excited by a localized source. When the excitation is by a well-focused distributed aperture, discrete or continuous superposition of point sources may be inconvenient. It is then desirable to seek here as well a collective alternative that facilitates the calculation for such beam type fields.

Thus, the objective of this fundamental study is the construction of a new theory of propagation, transmission and scattering that has broad implications for a general class of electromagnetic and other wave problems.

2. APPROACH

The proposed technique for the point source excitation problem is to apply partial Poisson summation to a group of wave events. The Poisson formula expresses these events rigorously in terms of their Fourier transforms, and of truncation (remainder) terms. The Fourier transforms provide wave events of another type whereas the remainder terms yield collectively weighted wave events of the original type. The relative importance of the Fourier transformed or collectively weighted original wave types determines the nature of the final field representation. This novel and general procedure contains as a special case the previously developed hybrid ray-mode formulation. The general procedure is also to be explored within the framework of rigorous contour integral representations when these are available.

When the source is an aperture with tapered illumination which generates a well-collimated beam, simulation of the source field distribu-

tion by a local evanescent plane wave profile permits collective and direct tracking of the field via the theory of complex rays, without the need for an integration over the plane wave spectral components. For such excitations, this collective approach to the source problem is to be combined with that described above for the propagation or transmission problem.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contribution can be understood more clearly.

The theory described in the next section is being applied to a number of propagation, transmission and scattering problems to confirm its versatility and to test its accuracy. The accuracy tests are performed by comparisons with solutions for canonical problems that can be generated by other means: numerical evaluation of vigorous series or integral representations, when available, or purely numerical solutions for (non-separable) problems that do not yield analytic solutions in explicit form. These applications are summarized below. All of the studies performed so far are for two-dimensional configurations in order to eliminate complexities in detail that are not essential in establishing the basic validity of the collective approach and of the ray-mode equivalent. A comprehensive survey was presented as an invited paper at the plenary session at a technical conference.[21]

1) Time-Harmonic Fields

(a) Singularity-free field tracking

One class of problems deals with repairing the deficiencies in an asymptotic ray theory (ART) analysis of high-frequency propagation in a highly overmoded waveguide that has an inhomogeneous refractive index profile in the transverse direction x. If the refractive index decreases monotonically from the perfectly reflecting top boundary to the bottom boundary, which separates the waveguide from an exterior semi-infinite medium with (lower) constant refractive index, one encounters the following categories of rays originating at the source: a) rays which are continuously refracted toward the top without encountering the bottom; b) rays which encounter top and bottom but are totally reflected at the bottom; c) rays which encounter top and bottom but are refracted into the exterior. ART fails (it predicts infinite fields) in the following transition regions: near the caustics formed by the surface guided rays in category a); near the bottom-glancing ray that separates categories a) and b); near the critically incident ray that separates categories b) and c). By the ray-mode equivalent, spectral intervals surrounding these transitional rays are filled with modes. The theory has been developed and its numerical implementation is in progress for a model waveguide with exponentially varying index profile. Preliminary results have been presented at a technical conference.

(b) Waveguides with longitudinal variation

When the refractive index and (or) the waveguide boundaries change along the guiding direction z, the resulting generally non-separable boundary value problem can be attacked by coupled mode theory provided that one invokes discretizing approximations (by inserting a false boundary in the exterior medium) to eliminate the continuous mode spectrum. The modes in question are those pertaining to the local z-independent environment. To reduce the intermode coupling, it would be useful to search for a more general class of local modes that adapts better to the changing environment than those in the plane parallel model. Such "more adaptable" modes are the "adiabatic" modes which, to the lowest order in the non-separability parameter, propagate without coupling to other adiabatic modes. Thus, they form an uncoupled system in waveguides with weak longitudinal variation. By two different approaches, one based on collective treatment of ray spectra, the other on scaling of mode spectra, we have developed a new theory for tracking these modes, with inclusion of the continuous spectrum, along the waveguide and also through the cutoff region if the configuration is such that an originally trapped adiabatic mode gets converted into a radiating (leaky) mode. These two approaches, which can be shown to yield the same solution, are discussed in more detail in Section 4. They have been published,[14,15] and a paper has been presented at a technical conference.[23]

(c) Complex ray modeling of distributed aperture fields

When the source of radiation is distributed over a large aperture instead of being localized, conventional ray tracing techniques become cumbersome and even invalid when the aperture field is tapered so as to generate a collimated beam. To circumvent the need for integration over the large aperture surface to pass from the aperture to the far zone, we have proposed to simulate the aperture field by Gaussian-like distributions that can be generated by a source point in complex coordinate space or, more generally, by an initial distribution of complex rays. The resulting fields can then be traced from the analytically continued aperture to the far zone without integration over the aperture surface. Work on this complex ray simulation of various taper profiles is in progress. For the Gaussian-like fields generated by a complex source point, the tracing of the beam from the aperture through a cylindrical shell radome to the far zone has already been accomplished. This complex ray analysis of radome-covered antennas was reported at a technical conference.[24]

2) Transient Fields

(a) Hybrid wavefront-resonance formulation of transient scattering

By extending the concept of rays, modes and their equivalents into the transient domain, we have developed a new theory of transient scattering that combines wavefronts (rays) and complex resonances (modes) in a self-consistent framework for efficient analysis of a target response at all observation times. On the prototype problem of impulse scattering by a circular cylinder, we have shown how the cumulative effect of wavefront arrivals generates the complex resonances of the

singularity expansion method (SEM) and, conversely, how the cumulative effect of resonances generates wavefronts. This new theory, which has been published,[16] is discussed in more detail in Section 4.

(b) Spectral theory of transient fields

Progress has also been made on reformulating transient propagation and scattering by inverting the conventional procedure, which first solves for a source-excited field in terms of a wavenumber spectral integral in the frequency domain and thereafter, by Fourier inversion, passes to the time domain. Certain advantages accrue by performing the frequency integration before the integration over the wavenumber spectrum. We have endeavored to generalize this approach, introduced by Chapman[17] for seismological problems, by viewing it within the context of weakly dispersive propagation processes. Our generalization of the Chapman procedure, which is constrained to real frequencies and spectral wavenumbers, involves analytic functions in the complex frequency and spectral wavenumber planes and therefore avoids Chapman's use of the theory of distributions and of Hilbert transforms. These ideas have been presented at a technical conference[25] and are being developed further.

4. STATE OF THE ART AND PROGRESS DETAILS

A. Introduction

Many electromagnetic propagation environments, whether natural or man-made, are so complicated that direct solution of the field equations to determine signal characteristics is beyond the scope of present analytical and computational capabilities. A common procedure has been to decompose the complicated, intractable propagation process into a sequence of simpler tractable events, and to synthesize the original problem by superposition of, and interaction among, the simpler constituents. Examples are provided by propagation in tropospheric layers, in the earth-ionosphere waveguide, along the surface of the earth, in optical fibers, etc., where the presence of transverse medium inhomogeneities or boundaries causes multiple reflections or ducting of the source-excited radiation. Here, the simpler problem would be the radiated field in the absence of boundaries, with the influence of the latter accounted for by multipath effects. In such a guiding environment, the field may alternatively be expressed in terms of discrete and continuous guided modes, but many of these are required at high frequencies when the duct width is large compared to the local wavelength, and "non-canonical" environments may introduce mode coupling. A somewhat different viewpoint prevails for propagation through, rather than along, stratified media as, for example, in vertical ionospheric sounding or in propagation through layered dielectric radome covers for aperture antennas. Here, guiding effects are generally unimportant but emphasis is placed on multiple reflections due to successive interfaces or medium gradients along the transmission path. Moreover, with respect to the excitation, different propagation phenomena are relevant when the incident field is due to a localized source and therefore has a broad radiation pattern, or due to a distributed aperture that generates a well-collimated beam.

For the class of problems alluded to above, superposition (by multiple reflection) of individually tractable propagation events constitutes one of the principal tools of analysis. At high frequencies, these events may be approximated as ray fields which undergo reflection, refraction and(or) diffraction on their path from a source at S to an observer at P. While ray theory provides a fundamental view of the propagation process by tracking local plane wave fields emanating at the source, such tracking becomes cumbersome when many ray paths exist between S and P. It would therefore be desirable to deal with multipath effects collectively. In guided propagation along a refracting channel, rays may form caustics (convergence or focusing zones of enhanced field strength) where simple ray theory fails. When these caustics are sufficiently distinct, one may correct ray theory by uniform asymptotic transition functions, but situations arise for rays with many reflections where an accumulation of caustics makes such corrections impractical and even impossible. Here, again, a collective alternative to multiple ray reflections is desirable or actually necessary. In transmission through a layered medium, an attractive alternative to tracking individual internal interface reflections and refractions would be to deal with these collectively by defining a "local slab" transmission or reflection coefficient.

The modal approach to ducted propagation is beset with similar difficulties when the required number of modes is large. It would be useful to express the interference properties of clusters of modes collectively in terms of simpler events. Moreover, asymptotic mode fields in complicated refracting media possess caustics near which that simple description fails. Collective treatment of such mode clusters could alleviate this difficulty.

The above considerations pertain to excitation by a localized source with a broad radiation pattern. Many applications deal, however, with distributed aperture excitations that generate a well-collimated beam field. These latter excitations can be synthesized in terms of the point source fields either by simulated discrete superposition or by continuous integration over the aperture domain. Alternatively, the source field could be decomposed into a plane wave spectrum and the observed field recovered by spectral synthesis of the plane waves after they have traversed the propagation environment. Each of these approaches requires an additional superposition or integration, which one would like to avoid if the excitation is of the beam type. Phrased in another way, one would like to be able to deal collectively with the distributed source problem for focused aperture fields.

The preceding discussion has made evident the importance of collective treatment of mode or ray fields when many of these are required to synthesize the signal in a particular transmission or guided propagation channel, or when failures in approximate mode or ray theory make these descriptions inapplicable. Substantial progress in this direction has been made through the new hybrid ray-mode theory developed by us recently. It has been shown how clusters of ray fields excited by a localized source can rigorously be converted into clusters of guided mode fields plus a (usually small) remainder, and vice versa. The theory has been applied to a series of "canonical" problems involving guided electromagnetic propagation along concave surfaces (here, the guiding mechanism is provided by "whispering

gallery" effects), in tropospheric ducts, in plane parallel homogeneously filled waveguides, and in graded index waveguides. The theory has also had impact on other fields such as underwater acoustic propagation and, with generalization to time-dependent signals, the modeling of seismic events. Concern in these applications has been with the greater computational efficiency of the hybrid formulation, with the avoidance of singular regions in ray fields or mode fields by filling these regions with mode fields or ray fields, respectively, and also with the penetrating physical insight of the propagation mechanism provided by the hybrid method. These accomplishments have been documented in a comprehensive series of publications.[1-11]

With respect to the source problem, it has been shown previously that well-collimated fields excited by tapered aperture illuminations can be analyzed effectively by use of complex rays and evanescent waves, without the need for aperture integrations or plane wave spectral synthesis. The presently proposed program deals in a general combined fashion with the collective description of a group of propagation events as well as a class of focused aperture excitations. The former was motivated by the hybrid ray-mode formulation and contains this formulation as a special case, whereas the latter is based on our previously performed evanescent wave and complex ray studies. We believe that application of the collective point of view to wave phenomena as well as distributed sources can make substantial impact on complicated propagation problems. The analytical setting for the collective wave treatment is provided, in general, by the Poisson summation technique, which is described below.

B. Foundation of the Collective Treatment

The collective method provides an inherently rigorous scheme for combining ray fields and mode fields (with inclusion of a remainder, when necessary) in well-defined proportions. For laterally homogeneous but vertically stratified media, the method can be derived either from an initial plane wave spectral representation of the Green's function or by application of partial Poisson summation. Via the former route, one decomposes the integrand in the plane wave spectral formulation in various ways so as to generate via contour deformation generalized ray integrals, or normal and leaky mode contributions (including continuous spectra). When only some of the ray integral or modal contributions are retained, the effect of the omitted contributions can be expressed in terms of mode and ray fields, respectively, plus remainder terms. The necessary manipulations in the complex wavenumber plane to achieve this hybrid representation have been illustrated in the various publications mentioned earlier. Alternatively, one may proceed by applying Poisson summation to a selected group of ray or mode fields, without starting first from an integral representation of the Green's function. Because the Poisson summation route is generally applicable to determine the collective behavior of any sequence of terms, and also because it highlights the physical mechanisms, it is discussed briefly in the following.

We begin with the Poisson identity for a group of terms f_q:[12]

$$\sum_{q=Q_1}^{Q_2} f_q = \frac{1}{2} f_{Q_1} + \frac{1}{2} f_{Q_2} + \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \int_{2\pi Q_1}^{2\pi Q_2} f(\tau) e^{ip\tau} d\tau \qquad (1)$$

In (1), $f(\tau)$ is obtained from f_q on replacing the discrete index q by $(\tau/2\pi)$, with τ representing a continuous variable. Let us assume that each f_q may possess in general a slowly varying amplitude A_q and a rapidly varying phase ψ_q so that

$$f_q = A_q \exp(i \psi_q) \qquad (2)$$

Alternatively, f_q could be defined in terms of an integral (for example, over a spectrum of plane waves), with the index q occurring in the integrand. In that event, that integration remains in the corresponding transition to $f(\tau)$ on the right-hand side of (1). It should be noted that f_q in wave problems also depends on other parameters, for example, source and observation points and medium parameters, and that the analytic behavior of f_q and $f(\tau)$ may be strongly affected by these. Evidently, the Poisson sum formula expresses the events f_q collectively in terms of their Fourier transforms with respect to the "smeared out" summation index, plus truncation effects depicted by (1/2) of the contributions from the first and last element.

To treat the τ-integral in (1), it is convenient to perform an asymptotic evaluation by the method of saddle points. With $f(\tau)$ expressed as in (2), contributions will arise from possible saddle points and from the endpoints. The saddle points τ_p are determined from the stationarity of the composite phase $\bar{\psi} = \psi + p\tau$:

$$d\bar{\psi}/d\tau = 0 \quad \text{at} \quad \tau_p \qquad (3)$$

and only those solutions of (3) that lie on or near the integration path between τ_1 and τ_2 will be relevant. Here, $\tau_{1,2} \equiv 2\pi Q_{1,2}$. If the saddle points are isolated from one another and the endpoints, the simple conventional saddle point formula applies.[13] If two or more saddle points are in close proximity, if one or more saddle points lie near an endpoint, or if a singularity is nearby, one must employ the appropriate uniform evaluation of the integral.[13] Relevant saddle points, if any, (i.e., those that yield dominant contributions to the leading asymptotic order) will usually occur only for a restricted range of p. Therefore, the infinite sum over p is effectively truncated to a finite number of terms. If f_q expresses a ray or ray-like event, the saddle point contribution furnishes a modal or mode-like event, and vice versa, since these two events are related by Fourier transformation. In particular, if f_q is a ray field, and the phase ψ_q characterizes the trajectories of a q-times reflected ray, the saddle point condition (3) defines the resonance (eigenvalue) equation for one or more guided modes distinguished by the mode index p. Conversely, if f_q is one of the congruences of a modal field of index q, and ψ_q characterizes the

modal congruence phase, the saddle point condition (3) defines the path equation for a p-times reflected ray passing from source to observer. Note that in either case, the mode or ray parameter τ_p lies between the limiting values $\tau_1 < \tau_p < \tau_2$ that define the extent of the original ray or mode group, respectively. Thus, the equivalent modes or rays occupy wholly or in part the same interval occupied by the original ray or mode group.

Endpoint contributions to the τ-integral in (1) are always present. If no saddle points or singularities are close by, the upper endpoint furnishes the leading asymptotic contribution

$$\int^{2\pi Q_2} f(\tau) e^{ip\tau} d\tau \sim \frac{f(2\pi Q_2) e^{i2\pi p Q_2}}{i\, d\bar{\psi}/d\tau \Big|_{2\pi Q_2}} = (-1)\left[\frac{d\psi}{d\tau} + p\right]^{-1}_{2\pi Q_2} f_{Q_2} \qquad (4)$$

For the lower endpoint, Q_2 is replaced by Q_1 and the result multiplied by (-1). Thus, the endpoints do not convert one wave type into another but they produce, with weighted amplitude, the limiting elements in the original group. Since the sum over p can be expressed in closed form as a cotangent function,

$$\sum_{p=-\infty}^{\infty} \frac{1}{\upsilon - p} = \pi \cot(\upsilon\pi) \qquad (5)$$

it follows that the endpoint contributions from the Poisson integral express the truncation of events in the original group as a single collective event associated with the first and last element:

$$\frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \left[\int^{2\pi Q_2} + \int_{2\pi Q_1}\right] d\tau \sim \frac{1}{2} \Delta_{Q_2} f_{Q_2} - \frac{1}{2} \Delta_{Q_1} f_{Q_1},$$

$$\Delta_{Q_{1,2}} = i \cot(\pi\, d\psi/d\tau)_{\tau_{\tau_{1,2}}} \qquad (6)$$

When all of these results are combined, one obtains the following transformation of (1):

$$\sum_{q=Q_1}^{Q_2} f_q \sim \sum_p F_p + \frac{1}{2}(1 + \Delta_{Q_2}) f_{Q_2} + \frac{1}{2}(1 - \Delta_{Q_1}) f_{Q_1} \qquad (7)$$

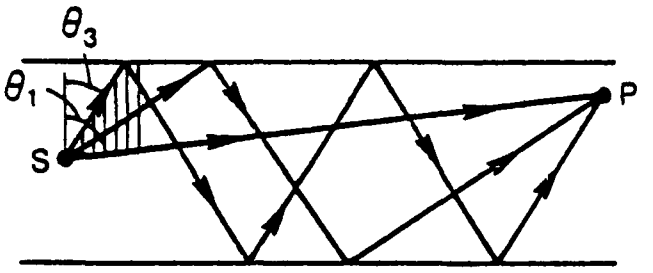
where F_p represents the contribution from relevant saddle points (isolated or in clusters) in the Poisson integral. Each F_p is a collective alternative of f_q in that they are Fourier transforms of one another; for example, if f_q belongs to a family of rays, F_p belongs to a family of

modes, and vice versa. On the other hand, the endpoints of the Poisson integral yield a collective description of the set $\{f_q\}$ in terms of the first and last elements, appropriately weighted, of the same set. Thus, if $\{f_q\}$ is a set of rays, then the endpoint contributions express its truncation as two equivalent collective rays. Similarly, a group of modes gives rise, due to truncation, to two equivalent collective modes. To the leading asymptotic order, the saddle point contributions dominate over the endpoints. Therefore, if relevant saddle points exist, the collective behavior of the set $\{f_q\}$ is predominantly characterized by the alternative set $\{F_p\}$, with the collective elements $f_{Q_{1,2}}$ providing some fine tuning.

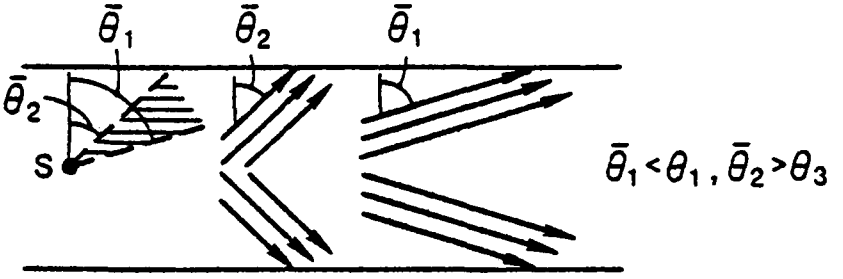
When saddle points or singularities approach the endpoints $\tau_{1,2}$ of the integration interval, the simple formula in (6) must be modified. Evidently, such a circumstance defines a transition region, where one of the elements F_p may escape from, or enter, the group. The appropriate uniform functions (Fresnel integrals)[13] ensure the smooth behavior of the total field as this transition takes place. When a saddle point coincides with an endpoint, the asymptotic result furnishes one-half of the saddle point contribution and there is then no endpoint contribution.

The results described above have a cogent physical interpretation which can be illustrated on a simple example. Consider a single homogeneous layer wherein a source S transmits a field to an observer at P. We concentrate on wave phenomena that leave the source in the upward direction and arrive at P from below. Out of the complete set of direct and multiple reflected rays f_q, we have retained only the direct ray and the first two reflected rays with angles $\theta_{1,2,3}$ at S; i.e., $q = 1,2,3$ (Figure 1(a)). These rays occupy the vertically shaded angular interval. The normal modes in the layer can be decomposed into upgoing and downgoing plane waves at the characteristic mode angles $\bar{\theta}_p$. In Fig. 1(b), it is assumed that only two adjacent modes, with characteristic angles $\bar{\theta}_1$ and $\bar{\theta}_2$, fall within the angular interval $\theta_3 < \bar{\theta}_p < \theta_1$; i.e., $\tau_{1,2}$ are the only saddle points on the integration interval in (1). Since the "mode volume" between $\bar{\theta}_1$ and $\bar{\theta}_2$ does not fill the initial "ray volume" between θ_1 and θ_3, there are intervals empty of modes (Figure 1(c)). Their effect is accounted for by the collective ray portions $(1/2)\Delta_{Q_{1,2}} f_{Q_{1,2}}$ in (7). If the mode volume fills the entire ray volume, then $\Delta_{Q_{1,2}}$ disappears and the first and last modes (here the only modes) contribute with half strength. The resulting reduction of (7) then exhibits complete symmetry between the group of rays and the group of modes, with the first and last element of each halved to account for truncation. Analogous considerations apply when an initial group of modes is to be converted into rays.

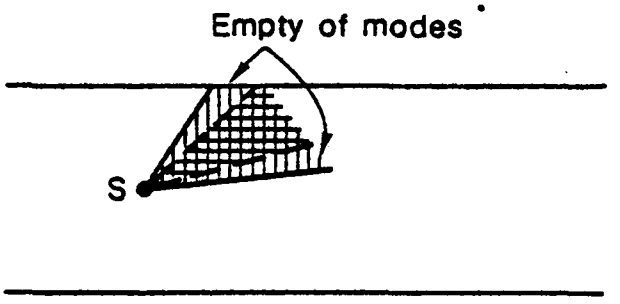
Thus, any group of rays can be converted into a well-defined group of modes and vice versa, with inclusion of collective rays or modes, when needed. This is the essence of the hybrid approach (see Figure 2). The asymptotic considerations above clarify the physical mechanism, but the formula in (1) is exact and can be used directly for numerical evaluation.



(a) group of rays with departure angles θ_α



(b) corresponding group of modes with characteristic angles $\bar{\theta}_p$



(c) composite domain, with empty regions

Fig. 1. Ray-mode equivalence in a homogeneous layer. Three rays in the depicted category (upgoing at S and at P) are to be converted into modes. At a fixed frequency, the characteristic mode angles are constant but the ray-mode combination is range-dependent since the ray paths change with relative positions of S and P. In the time-dependent regime, the hybrid combination changes even for fixed S and P (fixed ray angles) since the characteristic mode angles now vary with frequency.

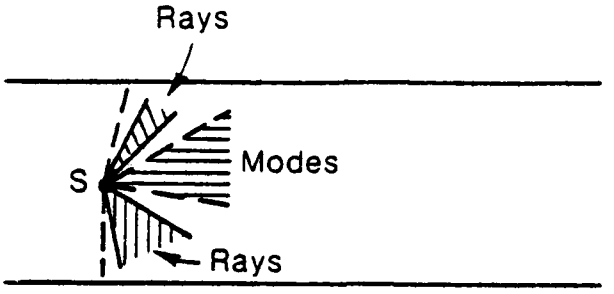


Fig. 2 Schematic representation of hybrid scheme. Alternate angular intervals at the source are filled with rays (vertical shading) or modes (horizontal shading). The effect of empty intervals (blank) is accounted for by collective rays or modes.

C. Previous Progress

An inhomogeneous waveguide with monotonically decreasing refractive index can trap rays and modes near the maximum index surface. The ray fields form caustics which may pile up at long distances from the source. The hybrid ray-mode approach can be employed to avoid inclusion of uniform asymptotic corrections of ray theory as the observer approaches the caustics. This is demonstrated in an example in Figure 3.[11]

The configuration of caustics and endpoints for various ray species is shown in Fig. 3(a), while the magnitude of the Green's function at fixed range and variable depth is shown in Fig. 3(b). Here, the reference (exact) solution (solid curve) is generated from the rigorous mode series (32 modes) and numerical evaluation of a continuous spectrum integral. The number of rays and modes belonging to ray species $j = 1,\ldots, 4$ in the hybrid formulation (dashed curve) is indicated on the figure. These ray species are ordered according to upward and downward departure at the source and to arrival at the observation point from above and below. For example, at $x/a = 0.05$, the hybrid representation involves 2 rays of species 1 plus 1 mode, 1 ray of species 2 plus 8 modes, 3 rays of species 3 plus 1 mode and 2 rays of species 4 plus 1 mode. A ray representation alone (with caustic correction) is utilized beyond $x/a = 0.115$; it is found to coincide completely with the reference solution. In this example, only ray fields with one or two reflections may be identified; this corresponds to $(x/a) > 0.1$. The ray method is inapplicable for rays with more than two reflections because of the pile up of corresponding caustics near the boundary. The locations of the caustics and cusps for the once and twice reflected rays, as inferred from Fig. 3(a) with $z/a = 1.92$ and variable x/a, are indicated below the abscissa. Note that the caustic corrections required in the ray formulation are avoided by the properly selected hybrid mix.

This study has been carried further by terminating the vertically inhomogeneous profile at $x = b$ by a semi-infinite medium with constant refractive index $n_2 < n(b)$. There now occur new transition regions in

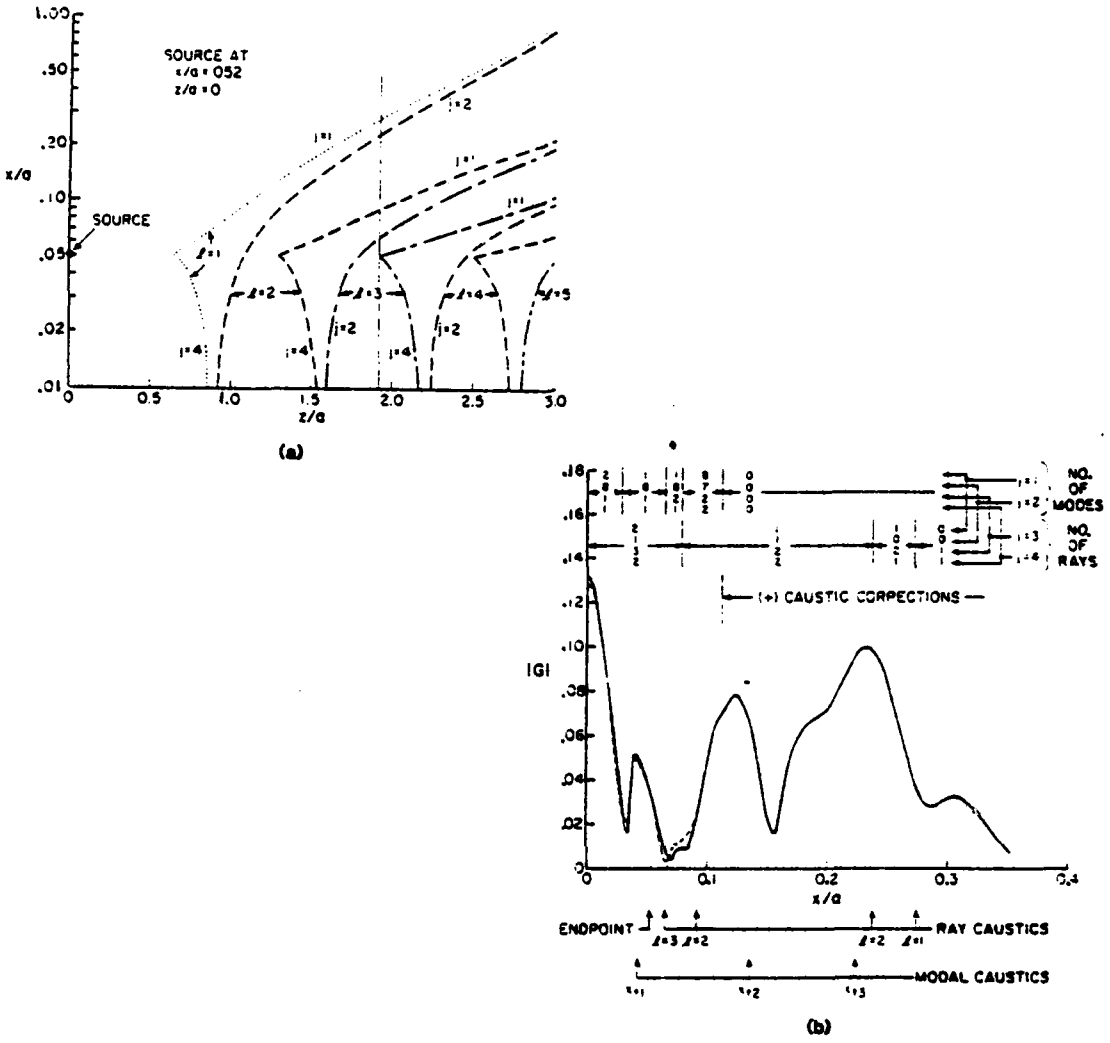


Fig. 3 Surface duct with exponential refractive index $n^2(x) = \exp(-2x/a)$. A time-harmonic line source with normalized frequency ka = 100 is located at the normalized coordinates x/a = 0.052, z/a = 0. (a) Caustics associated with rays of species j = 1,...,4, which have undergone ℓ reflections on the boundary x = 0. (b) Variation of signal strength |G| for observation points located at a range z/a = 1.92, with variable depth x/a. Solid curve —— exact solution; dashed curve --- hybrid ray-mode solution; the number of rays and modes for each ray species is indicated on the figure.

the vicinity of the glancing ray and the reflected ray excited by a ray incident at the critical angle. The failure of ray theory in these transitional domains has been corrected by replacing the angular spectra surrounding these rays by a bundle of modes plus collective remainders. These remainders can be phrased either as collective rays or collective modes (see Sec. B). Numerical results are being computed to assess which of these options is preferable from the standpoint of computational simplicity, efficiency and accuracy. Direct numerical integration of the exact integrals in the formulation is to serve as a reference solution for comparison.

D. Recent Progress

1) Transient fields

For transient fields, a new theory of scattering has emerged by examining within the context of the collective treatment the canonical problem of diffraction of a pulsed incident signal by a perfectly conducting cylinder. Here, alternative descriptions of the scattering process can be developed either in terms of wavefronts traveling along ray trajectories or in terms of oscillatory fields (resonances) for the object as a whole. The wavefronts represent local discriminants because they carry information about the object gathered essentially along the direct, reflected and refracted ray paths only. Because the ray paths generally change with the location of source and receiver, this information is strongly source and receiver dependent (aspect dependent). The resonances (complex for exterior problems) represent global discriminants because they are self-consistent damped oscillations determined by the entire object surface. They are independent of source and receiver (aspect independent), although their strengths of excitation and reception are influenced by the source-receiver arrangement.

For the cylinder problem, as for any smooth convex scatterer, the wavefronts involve incident, reflected and (diffracted) creeping wave constituents. They are useful at early observation times when, due to causality, only a small number of them arrives at the observer. Of special interest are the creeping waves, which reach the observer after successive circumnavigations of the cylinder. These successive arrivals can be readily resolved at early times, but not so at later times when the memory of earlier arrivals blurs the contribution from the later ones; moreover, the summation of an increasing number of arrivals makes this description inconvenient. The resonances, on the other hand, which synthesize the transient field as a sum of damped oscillations (the complex frequency pole contribution in the singularity expansion method (SEM)),[18] are inconvenient at early times because many of these oscillatory (global) constituents are then required to generate a zero field ahead of the wavefront and an abruptly changing non-zero field behind it (Fig. 4(a)). This leads to convergence problems and the need for including an additional entire function in the frequency domain representation. At later times, however, only a few of the most weakly damped resonances survive and yield an effective transient field description (Fig. 4(b). Within the framework of our theory, wavefront sums are shown to be expressible collectively as resonances, and resonance sums collectively as wavefronts. This leads to a new theory of transient scattering wherein some wavefronts and some resonances are retained by self-consistent combination so as to exploit the advantages

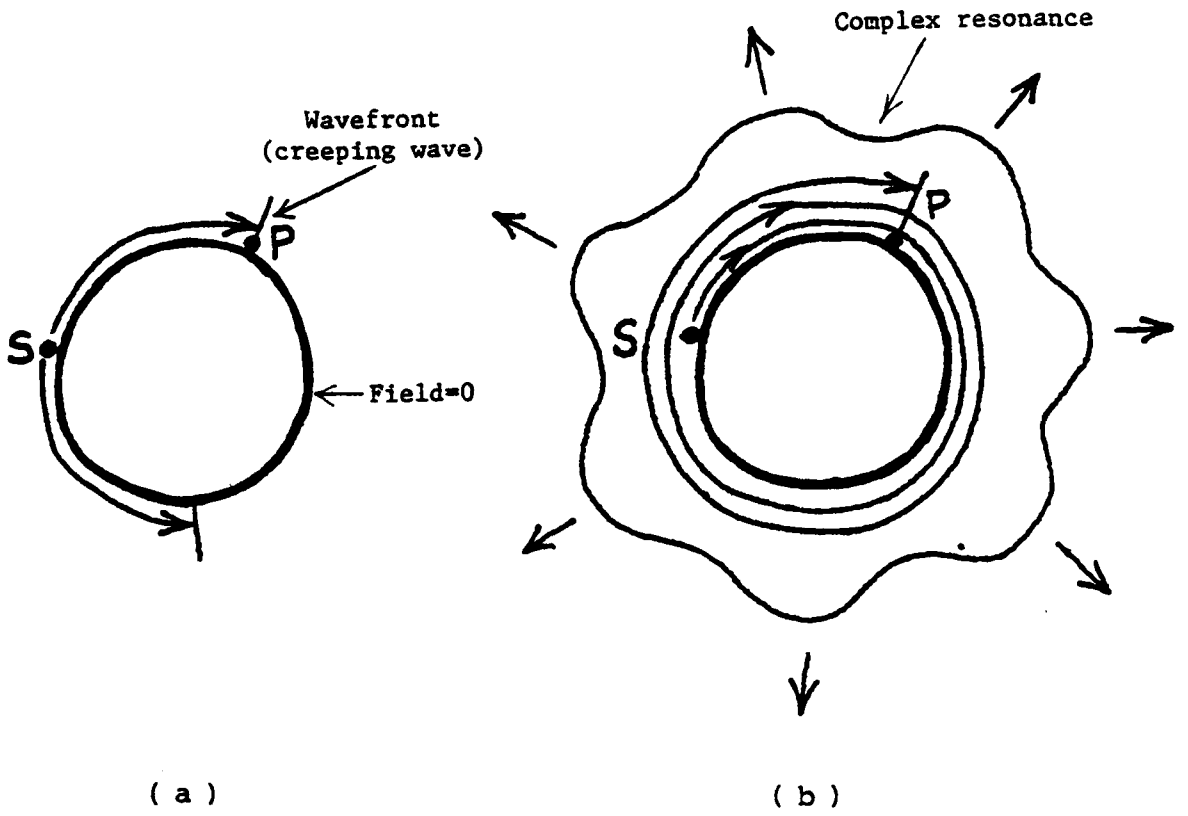


Fig. 4 Phenomenology for transient scattering by a cylindrical object. (S = Source, P = Observer) (a) early times: one wavefront corresponds to many resonances, (b) late times: many passages of wavefront correspond to few resonances (only the clockwise revolutions are shown).

of wavefronts at early times and of resonances at later times (Fig. 5). While it has previously been recognized by semi-heuristic arguments and by computer experiment[19] that sums of resonances generate traveling wave disturbances, the vigorous formulation of the wavefront-resonance equivalent and of their incorporation into a hybrid scheme is new. A publication describing these results has appeared.[16]

2) Range-dependent waveguides

Substantial effort has also been devoted to open waveguides with (non-separable) longitudinal variation. Here, the conventional approach has been via coupled mode theory. In that theory, one generally ignores coupling to the continuous radiation spectrum, although efforts have been made to discretize the spectrum by inserting artificial boundaries, including some loss, and then attempting to have the complex modes of this huge multimode waveguide describe the relevant wave process in terms of discrete spectra only. When, as in many physical problems, the longitudinal dependence is sufficiently slow over an interval of length equal to the local wavelength, simpler approximations involving adiabatic techniques can be applied. An adiabatic mode alters its character by smoothly adapting to the changing environment without, to the lowest order of approximation, coupling to other adiabatic modes. However, this simple description fails in transition regions where (in a narrowing waveguide, for example) an initially trapped mode encounters cutoff and is transformed into a radiating mode. The failure of adiabatic mode theory in the transition region has hampered its application to this important class of events. We have addressed the transition problem by two methods motivated by the collective approach to guided mode type and ray type events. The collective treatment in this general environment is based on the definition of an adiabatic invariant that ensures that the essential features of the collective process are retained intact.

a) Spectral scaling

By the first method, a plane wave spectral representation of the waveguide Green's function is scaled so that the local plane waves synthesizing a local (adiabatic) mode adjust their characteristic propagation angles or wavenumbers to satisfy the local transverse resonance condition for the modal eigenvalue. In a plane parallel dielectric waveguide, the modal resonance equation can be expressed generically as

$$\Phi_p\ (\zeta_p,\ a) = 2\pi p, \qquad p = \text{integer} \tag{8}$$

With ζ_p and "a" denoting the (constant) longitudinal wavenumber and waveguide height, respectively, Φ_p expresses the phase accumulation of a modal plane wave making one complete excursion between the boundaries. The equality in (8) ensures the phase coherence required by the p-th modal field. Since Φ_p = constant, this quantity can be regarded as the modal invariant. In a tapered waveguide, where a = a(z) changes with the longitudinal coordinate z, the invariant is changed to the form

$$\Phi_p\ [\zeta_p(z),\ a(z)] = 2\pi p \tag{9}$$

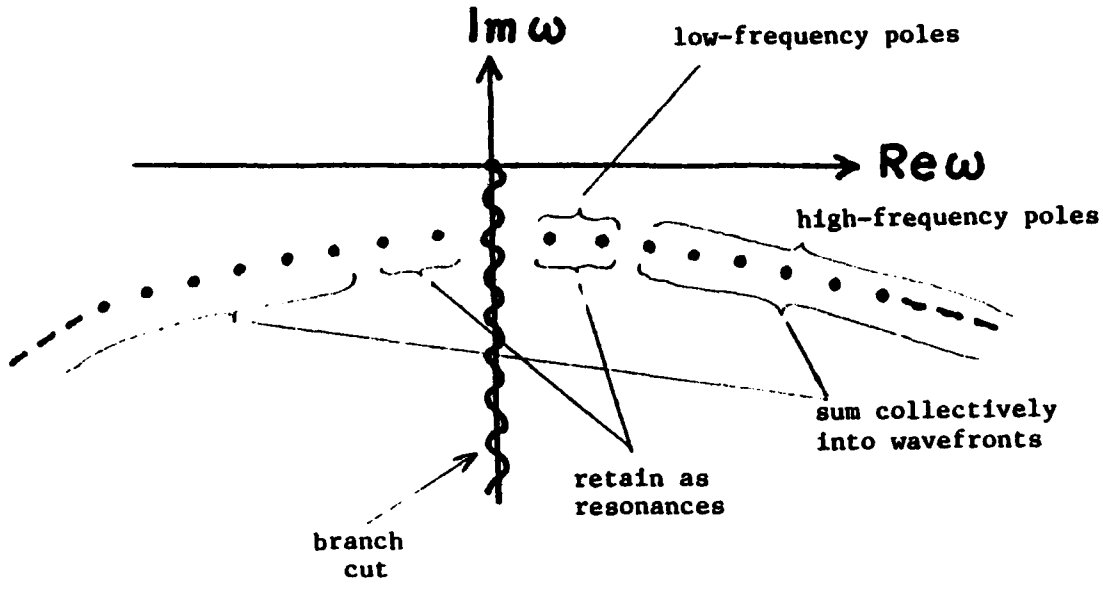


Fig. 5 Resonance (SEM) poles in complex frequency plane. Hybrid formulation:

Transient field $\Longrightarrow \sum f_p$ + $\sum F_q$ + (Wavefront or Resonance Remainder)

$\sum f_p$: some low-freq. resonances $\Rightarrow$ dominant at late times

$\sum F_q$: some wavefronts $\Rightarrow$ dominant at early times

thereby assuring its constancy by adjusting the transverse modal wavenumber ζ_p accordingly. In a plane wave spectral representation, which requires integration over ζ, it is thereby implied that one must scale not only the modal spectra ζ_p but all spectral values. Thus,

$$\Phi\,[\zeta(z_1),\, z_1] = \Phi\,[\zeta(z_2),\, z_2] = \text{constant}, \tag{10}$$

where z_1 and z_2 denote any two longitudinal coordinate points. The new concept expressed in (10) allows the coordinate separable spectral integral for the plane parallel waveguide to be generalized to weakly non-separable (for example, slowly tapered) waveguide configurations. The resulting integral contains all of the relevant propagation phenomena, including the transition of an adiabatic mode through cutoff. A numerical example for an ocean acoustic waveguide is presented in Fig. 6. Explicit formulas for the adiabatic modes and their transition behavior can be derived by asymptotic treatment of the spectral integral representation. Details may be found in a recent publication.[14]

(b) Ray-to-mode conversion

Ray fields in a longitudinally changing waveguide do not have to undergo spectral scaling since they describe ab initio a local propagation process. Ray fields can be constructed without constraints of separability, provided only that conditions change sufficiently slowly to validate local plane wave propagation along each trajectory defined by the ray equations. The ability to define local modes in a non-separable configuration can be tied to the ability to treat multiple reflected ray fields collectively by Poisson summation (Sec. 4B, Eq. (1)). For illustration, consider a wedge-shaped dielectric waveguide, with the upper boundary assumed to be inclined at an angle α with respect to the z-axis while the lower boundary remains parallel to z. The ray trajectories in the waveguide, along which the local plane waves synthesizing the ray fields are transported, can be constructed in this configuration conveniently in terms of image sources along a circle centered at the apex and passing through the source point. The phase $\psi_q(\zeta_q)$ for a q-times reflected ray can be constructed in the form

$$\psi_q(\zeta_q) = L_q(\zeta_q) + \hat{\Phi}_q(\zeta_q) \tag{11}$$

where

$$L_q(\zeta_q) = \sum_{\ell=1}^{q} (L_\ell + L'_q)\ , \quad \hat{\Phi}_q(\zeta_q) = \sum_{\ell=1}^{q} \frac{1}{k}\, \Phi_o(\theta_\ell) \tag{11a}$$

are the geometrical length of the ray path and the normalized composite phase from reflections at the waveguide boundaries, respectively. ζ_q is the longitudinal wavenumber of the local plane wave along the ray, referred to the z-coordinate, θ_ℓ is the ray reflection angle at the ℓ-th reflection, and Φ_o the phase of the ray reflection coefficient. Because the boundaries are non-parallel, θ_ℓ changes from one reflection to the next.

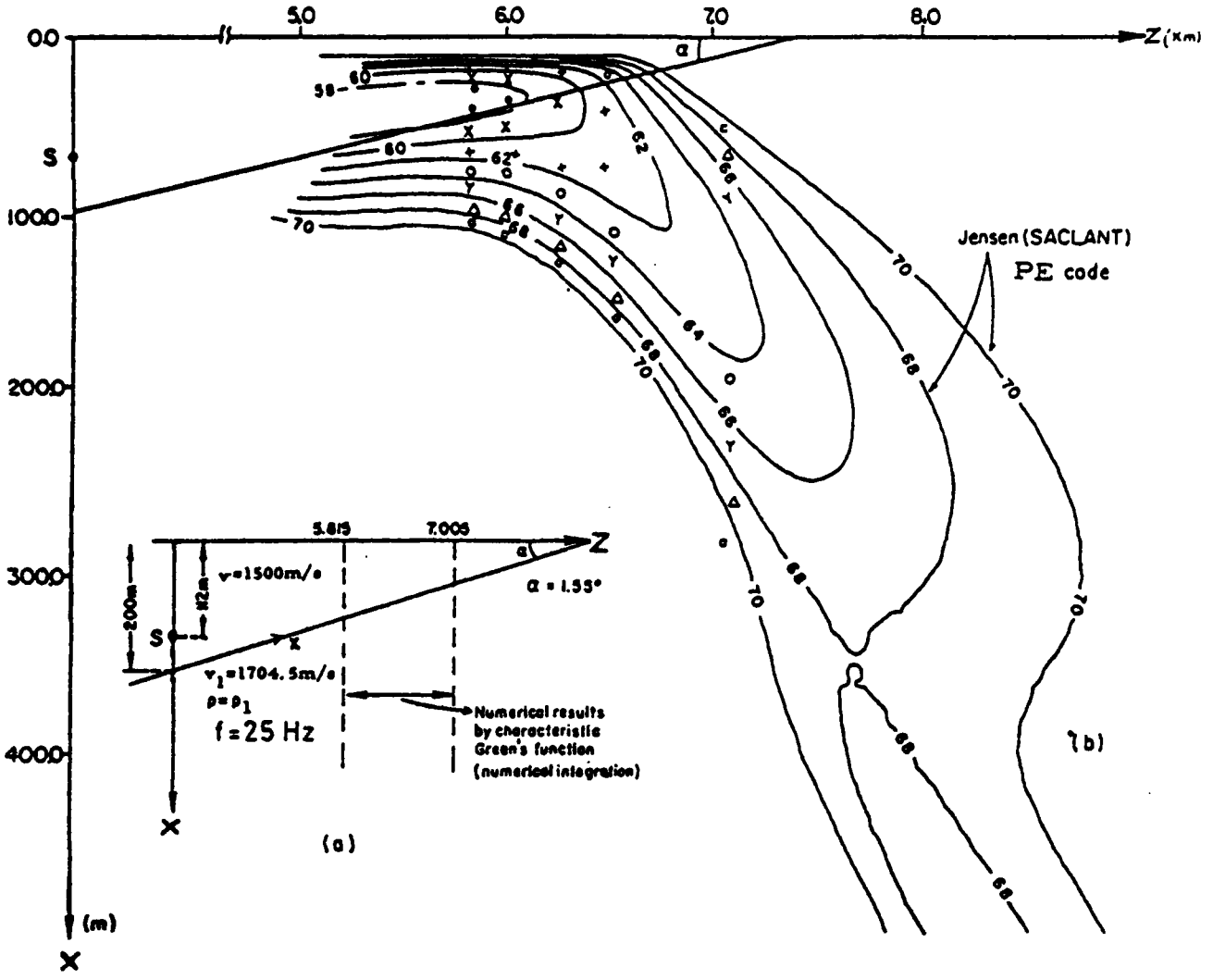


Fig. 6 Numerical results for pressure magnitude in an underwater acoustic channel with sloping bottom due to a line source at S. (a) Physical configuration. v and v_1 are sound speeds in the water and bottom, respectively, while ρ and ρ_1 are the corresponding acoustic densities. (b) Magnitude of acoustic pressure. The solid curves of equal acoustic pressure magnitude were calculated by the PE (parabolic equation) code. The dB level on each curve is relative to the field magnitude one meter away from the source. The points were obtained by numerical integration of our generalized spectral integral. □, Y, O, X, O, refer to 70, 68, 66, 64, 62, 60, 58 dB, respectively. Note that the horizontal and vertical scales are in kilometer and meters, respectively.

One may verify that the dependence of $L_q(\zeta_q)$ on q is such that the transition $q \to \tau/2\pi$ from discrete index q to continuous variable ζ, as required in the Poisson sum formula (Sec. 4B, Eq. (1)), can be performed directly. Likewise, the transformation of ℓ into a continuous variable can be carried out directly in each of the individual phase terms $\Phi_o(\theta_\ell) \to \Phi_o[\theta(\ell)]$.

However, this is not true for the composite phase $\hat{\Phi}_q(\zeta_q)$ since it depends on q through the upper limit on the sum in (11a). It is therefore necessary to perform a smoothing operation whereby the sum is replaced by a continuum, an integral. This difficulty is avoided in the plane parallel configuration were $\theta_1 = \theta_2 = \ldots = \theta_q$ (equal ray reflection angles) so that $k\hat{\Phi}_o(\zeta_q) = q\,\Phi_o(\zeta_q)$ in that special case. The smoothing for the tapered configuration is accomplished via the Euler-McLaurin formula

$$k\hat{\Phi}_q(\zeta_q) = \int_1^q \Phi_o[\zeta(\ell)]d\ell + \frac{1}{2}\,\Phi_o(\zeta_1) + \frac{1}{2}\,\Phi_o(\zeta_q) + E_q \qquad (12)$$

where the error term E_q involves successive derivatives that are small when α is small. The integral in (12) (as well as E_q) can be defined for $q \to \tau/2\pi$, thereby furnishing the continuous function $\hat{\Phi}[\zeta(\tau)]$. For sufficiently small α, E_q can be neglected, and the resulting $\hat{\Phi}[\zeta(\tau)]$ is then a lowest order approximation, in which the spectrum has been smeared out so as to express the discrete ray reflections as samples in a continuum.

One may now proceed with the Poisson sum formula, which transforms the ray fields E_q into modal fields F_p plus collective remainders (Sec. 4B). The saddle points in the τ spectral integral are found to furnish a modal resonance condition, which has exactly the form in (1) and thus, yields the local or adiabatic modes, each characterized by its modal invariant. Moreover, the spectral integral in the Poisson formula may be manipulated so that it agrees with the spectral integral referred to above. These considerations are discussed in detail in a recent publication.[15]

The two methods described above, which incorporate new approaches to propagation in open waveguides with weak longitudinal variation, are now being applied to more general configurations.

5. REFERENCES AND PUBLICATIONS

1. L>B> Felsen and A. Kamel, "Hybrid Ray-Mode Formulation of Parallel Plane Waveguide Green's Functions," IEEE Trans. on Antennas & Propag. AP-29, 637-649 (1981). Awarded Best Paper Prize for 1981 by IEEE Antennas and Propagation Society.

2. A. Kamel and L>B> Felsen, "Hybrid Ray-Mode Formulation of SH Motion in a Two-Layer Half Space," (date).

3. E. Niver, S.H> Cho and L>B> Felsen, "Rays and Modes in an Acoustic Channel with Exponential Velocity Profile," (date).

4. L>B> Felsen and T. Ishihara, "Hybrid Ray-Mode Formulation of Ducted Propagation," J. Acoust. Soc. Am. (March 1979).

5. T. Ishihara, L>B> Felsen and A. Green, "High Frequency Fields Excited by a Line Source Located on a Perfectly Conducting Concave Cylindrical Surface," IEEE Trans. on Antennas and Propag. AP-26, No. 6 (November 1978).

6. L>B> Felsen and T. Ishihara, "High Frequency Fields Excited by a Point Source on a Concave Perfectly Conducting Cylindrical Boundary," Radio Science, 24 No. 2, pp. 205-216 (March-April 1979).

7. T. Ishihara and L>B> Felsen, "High Frequency Fields Excited by a Line Source Located on a Concave Cylindrical Impedance Surface," IEEE Trans. on Antennas and Propag. AP-27, No. 2 (March 1979).

8. S>H> Cho, C>G> Migliora and L>B> Felsen, "Hybrid Ray-Mode Formulation of Tropospheric Propagation," AGARD Symposium on Special Topics in HF Propagation, Lisbon, Portugal, p. 11, 1-8, 15 (May 28-June 1, 1979).

9. L>B> Felsen, "Hybrid Ray-Mode Fields in Inhomogeneous Waveguides and Ducts," J. Acoust. Soc. Am., 69(2), 352-361 (February 1981).

10. E. Topuz and L>B> Felsen, "High-Frequency Electromagnetic Fields on Perfectly Conducting Concave Cylindrical Surfaces," IEEE Trans. on Antennas and Propag. AP-28, No. 6 (November 1980).

11. E. Topuz, E. Niver and L>B> Felsen, "Electromagnetic Fields Near a Concave Perfectly Conducting Cylindrical Surface," accepted for publication in IEEE Trans. on Antennas and Propag.

12. A. Papoulis, Signal Analysis, McGraw-Hill: New York, 1977, Sec. 3.3.

13. L>B> Felsen and N. Marcuvitz, Radiation and Scattering of Waves, Prentice Hall: Englewood Cliffs, NJ, Chapter 4 (1973).

14. A. Kamel and L.B. Felsen, "Spectral theory of sound propagation in an ocean channel with weakly sloping bottom," J. Acoust. Soc. Am. 73, pp. 1120-1130, 1983, 5. BASE Proposal (L.F. section).

15. J.M. Arnold and L.B. Felsen, "Rays and local modes in a wedge-shaped ocean," J. Acoust. Soc. Am. 73, pp. 1105-1119, 1983.

16. E. Heyman and L.B. Felsen, "Creeping waves and resonances in transient scattering by smooth convex objects," IEEE Transactions on Antennas and Propagation AP-31, pp. 426-437, 1983.

17. C>H> Chapman, "A New Method for Computing Synthetic Seismograms," Geophys. J>R> Astronom. Soc. 54 481-518 (1978).

18. C.E. Baum, "The singularity expansion method," Transient Electromagnetic Fields (L.B. Felsen, editor), Springer Verlag, New York, 1976.

19. H. Uberall and C.G. Gaunaurd, "The physical content of the singularity expansion method," Appl. Phys. Lett., 39, pp. 362-364, 1981.

20. L.B. Felsen, "Propagation and scattering of high-frequency and pulsed signals in the presence of a concave boundary," in New Procedures in Nondestructive Testing (Proceedings), Editor P. Hoeller, Springer Verlag, Heidelberg 1983. Paper #5.1.

21. L.B. Felsen, "Progressive and oscillatory representations for propagation and scattering," invited paper (plenary session), IEEE AP-S/URSI International Symposium, University of Houston, Houston, Texas, May 1983.

22. E. Niver, A. Kamel and L.B. Felsen, "More on transitional ray fields," Acoustical Society Meeting, Chicago, Illinois, November 1982.

23. J.M. Arnold and L.B. Felsen, "Ray invariants, plane wave spectra and adiabatic modes for range-dependent and shallow ocean acoustics," International Conference on Acoustics and the Sea-Bed, University of Bath, Bath, England, April 1983. Published in the Conference Proceedings, Bath University Press.

24. X. Gao and L.B. Felsen, "Complex ray analysis of radome-covered large aperture antennas," same conference.

25. L.B. Felsen and E. Heyman, "Transient analysis of weakly dispersive fields," same conference.

6. DoD AND OTHER INTERACTIONS

(a) Collaboration with Dr. J.M. Arnold, Univ. of Nottingham, Nottingham, England, on hybrid methods applied to tapered dielectric waveguide.

(b) Collaboration with Professor A.M. Scheggi, Institute for Electromagnetic Wave Propagation, Florence, Italy, on hybrid methods applied to optical fibers.

(c) Collaboration with NATO Underwater Sound Laboratory, La Spezia, Italy, on hybrid methods for propagation in ocean channels.

(d) Collaboration with Professor R. Zich, Technical Univ., Turin, Italy, and Professor F. Hasselmann, Catholic Univ., Rio de Janeiro, Brazil, on evanescent wave and complex ray methods for large reflector antennas.

(e) Collaboration with Professor E. Topuz, Department of Electrical Engineering, Technical University of Istanbul, Turkey, on application of hybrid methods to imaging in multi-mode optical waveguides.

Programs involving the basic ideas developed under JSEP have been supported under separate contract by the U>S> Army, by the U>S> Navy, and by the U>S> Geological Survey, and by the National Science Foundation.

E. HIGH POWER MICROWAVE-ATMOSPHERE INTERACTIONS

Professors N. Marcuvitz and W.T. Walter

Unit EM3-5

1. OBJECTIVE(S)

Study of high power microwave-matter interactions in the atmosphere in order to obtain a detailed space-time dependent description of the nonlinear and turbulent processes involved in the conversion of microwave energy into particle dynamic energy. Of applicational concern will be the determination of the increase of pressure and temperature that can be produced at different altitudes as a function of power level, pulse length, repetition rate, frequency, focusing configuration, etc.

This study continues the basic approaches developed earlier by us for understanding nonlinear and turbulent processes in wave-matter interactions. The primary concern here is with interactions between high power microwaves and the atmosphere.

2. APPROACH

The approach to the overall space-time description of high power microwave absorption in the atmosphere is based on a number of past and current theoretical and experimental wave-matter activities of faculty and graduate students working on nonlinear and turbulent wave propagation at both microwave and laser frequencies. Our approach is not to evolve a massive "code" for computation, but rather to concentrate on the analysis, in a modular sense, of those finite algorithmic models that affect the accuracy of existing procedures used in large overall codes, with particular emphasis on hydrodynamic processes that occur in the absorption of high power microsecond microwave pulses.

3. SUMMARY OF RECENT PROGRESS

1) Theoretical

This section briefly summarizes progress in the past period. We have explored analytically and numerically energy conversion processes associated with the interaction of a high power modulated microwave pulse and an atmosphere ionized by the pulse. The basic problem is that of a pulsed beam of finite spatial width and temporal extent incident obliquely on an atmosphere at power levels such that the plasma formed is nonlinear, turbulent, and space-time variable. The analysis of this problem is dependent on the derivation of suitable equations descriptive of the electromagnetic and particle dynamics, and capable of being manipulated into a form for efficient numerical analysis. To obtain computational efficiency we have employed quasiparticle methods to describe both the microwave fields and the collective processes excited in the plasma formed by the pulse; these methods average out fast space-time oscillatory behavior and retain only the relatively slow space-time variability of engineering interest.

A number of modular problems have been considered to obtain accuracy checks on our computational algorithms in a parametric range wherein exact results are known; this has been done in order to gain confidence in algorithmic accuracy for more extended ranges. For example, an extensive series of computational experiments has been rur in which a gaussian modulated microwave pulse of finite spatial and temporal width incident on a uniform electron plasma of finite and length collision frequency excites a known damped oscillatory electron-acoustic pulse propagating into the plasma; for sufficiently low carrier frequency the detailed space-time propagation of the pulse can be computed quite exactly and hence provides a check on the accuracy of the "macroscopic" quasiparticle description of the plasma pulse. A typical computational comparison of the exact time dependence and the quasiparticle approximation for a plasma pulse on propagating a finite distance into the plasma is shown in Fig. 1. Results for an extended range of parameters, both nonlinear and turbulent, have been and are being carried out in Ph.D. research by G. Kopcsay.[23]

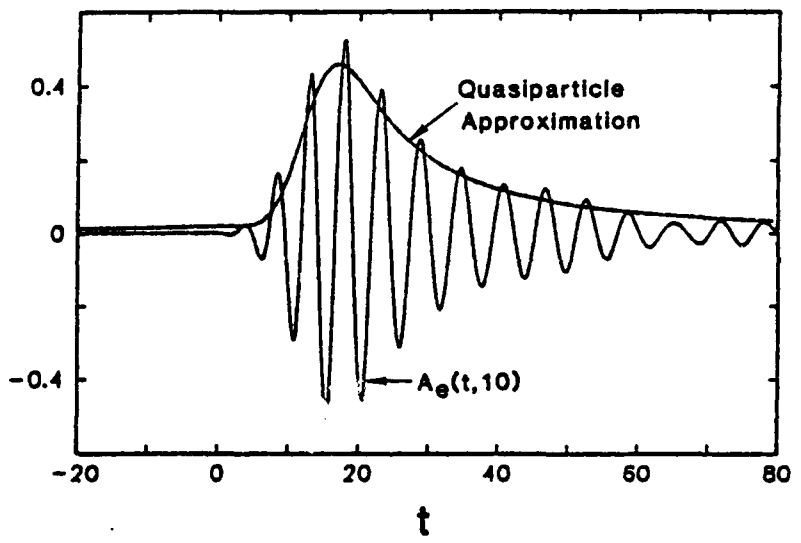


Fig. 1 Comparison of exact solution with quasiparticle approximation for the plasma pulse.

During the past period, some related Ph.D. research (S.Y. Lee), which is a spin-off of the above and also amenable to quasiparticle techniques, has resulted in publications that describe beam shifts and amplitude distortion on reflection and transmission through dielectric structures[24] and to frequency chirping and pulse distortion on propagation through dispersive dissipative media.[25] One of the difficulties with the first order quasiparticle procedure is that it is an approximation applicable in a range wherein beam size is large compared to wavelength but small compared to inhomogeneity scales in the propagating medium. Ordinary higher order approximations outside of this range are not

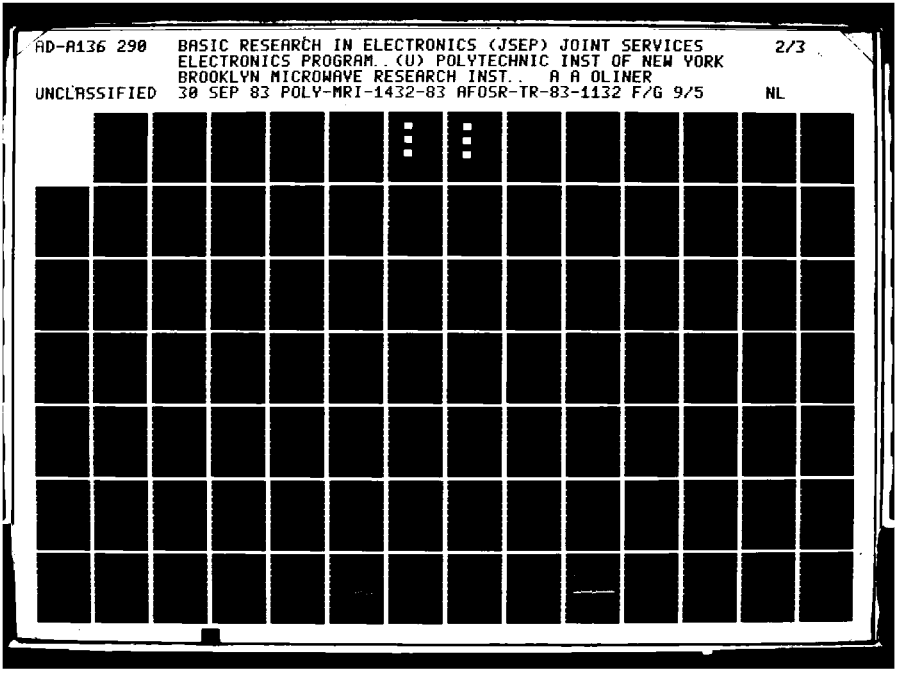
rapidly convergent and hence difficult to calculate; a continuing research effort by G. Wang is attempting to use regularization and renormalization group procedures to find rapidly convergent approximations valid for strong medium inhomogeneities. Previously reported Ph.D. work using renormalization and quasiparticle methods by D. Wu on beam propagation through turbulent ionospheric media[26] will be published in a few months. A paper on the quasiparticle method[27] was presented by N. Marcuvitz at an international URSI symposium on electromagnetics in Santiago de Compostela, Spain, in August 1983.

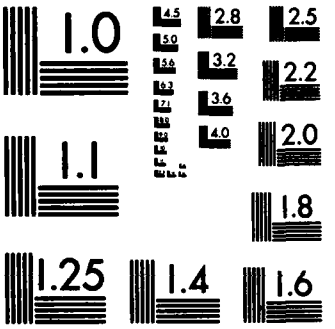
2) Experimental

The magnetron previously utilized in our experimental verification studies of our pulse propagation model had a peak power capability of 50 kW at 3 GHz and pulse widths of 80 or 400 nanoseconds. As previously reported, pulse shortening was observed within a pressure region of 2 to 20 torr around the Paschen minimum. We have noted that higher peak powers and longer pulses would be desirable to enlarge the pulse shortening region for comparison with the model. Since a 5795 magnetron was available in our storeroom with a 1 MW peak power capability, we have constructed a new high-power modulator with sufficient capability to drive megawatt magnetrons and with appropriate interlock protection both for the tube as well as for the protection of personnel. Upon completion of the modulator and testing of the magnetron, the 5795 tube was found to be gassy and unable to hold the high voltage. We wish to acknowledge the generous assistance of Raytheon Corporation. Through the assistance of Senior Engineer Richard Y. Clark we were able to obtain a 1448 magnetron from Raytheon. This tube was set for operation at 3.28 GHz ± 20 MHz and degaussed to provide a MW output pulsed when driven with 37 kV, 50 A peak pulses. The 1448 magnetron has now been installed on our S-band line. Thus far the magnetron has been operated up to 32 kV, and 3 μsec pulses have been detected at the side wall couplers within a 16 to 32 kV range of operation. A new rf peak power meter has been ordered and is expected to be delivered at the end of September. Data-taking with the higher-power magnetron is expected to begin in December.

3) Computational Facilities

In the past our theoretical modeling results could not be extended to long pulses because of insufficient addressable memory on the 16-bit MRI computer, a DEC PDP 11/60. During 1982 a network of four Apollo computational nodes, each with 1 megabyte of fast MOS memory and a 32-bit word structure, was added to the MRI Computational Facility. The computer model had been constructed using our scientific, interactive graphic language, IGL, which is a higher-level programming language particularly adapted to interactive computations and graphical display of results. Because it was written in assembly language, IGL is also quite fast but, of course, limited to the PDP 11 CPU. To make our interactive graphic language portable, we recently constructed a C-based version of IGL called TIPL, the interactive programming language. A number of additional improvements were also incorporated into TIPL including a more algebraic notation, a larger number of possible variables (scalars, vectors and arrays) with more mnemonic names, etc. Unfortunately the speed of TIPL is 2.5 to 5 times slower than IGL, largely due to the assembly to C language conversion. In


AD-A136 290 BASIC RESEARCH IN ELECTRONICS (JSEP) JOINT SERVICES 2/3
ELECTRONICS PROGRAM..(U) POLYTECHNIC INST OF NEW YORK
BROOKLYN MICROWAVE RESEARCH INST.. A A OLINER
UNCLASSIFIED 30 SEP 83 POLY-MRI-1432-83 AFOSR-TR-83-1132 F/G 9/5 NL



addition the speed of the Apollo computers turned out to be twice as slow as the PDP 11/60. Therefore, computer model programs now run ten times longer and cannot realistically be extended to longer pulses since many hours were already required for a single run on the PDP 11/60.

To overcome the twin problems of additional addressable space and increased speed, we utilized equipment funding from JSEP during 1983 to order a CSPI array processor on a Gould 32/6750 super mini-computer. The host Gould computer has a 32-bit word structure, 2 megabytes of fast memory and is expected to be about eight times faster than the PDP 11/60. In addition the CSPI 6410 array processor, which has 32K words of 32-bit program/data memory and 64K words of 64-bit data memory, can provide additional speed enhancements of up to 200 depending on the program. We are hoping for a speed enhancement of 10 to 20 which would then permit extension of our computer modeling results to longer pulses. The Gould 32/6750 and CSPI 6410 array processor arrived at the Polytechnic in September, 1983. We anticipate a three to four month effort to write a UNIX driver for the array processor. Modifications will also have to be made in the IGL computer program, so that initial results are expected in the spring of 1984.

4. STATE OF THE ART AND PROGRESS DETAILS

The current and anticipated availability of very-high-power microwave sources is leading to a renewal of interest in effects caused by dumping of large amounts of energy into atmospheric regions that have been broken down and ionized. The determination of the resulting enhancement of pressure, temperature, etc., of the atmospheric constituents requires a self-consistent space-time dependent description of the electromagnetic fields and of the electron, ion, and neutral particle dynamics within the interaction region. Of particular interest are effects arising from power levels above breakdown ($> 10^6$ watts/cm^2 at ground level) and pulse lengths in the microsecond range and longer. The equations descriptive of the interaction are the Maxwell equations together with kinetic and/or fluid dynamic equations for the particle constituents with coupling represented by "collision" terms for the different wave-particle, particle-particle, and wave-wave interactions involved. Researchers have employed a variety of descriptive models that require a mix of analytical assumptions, extrapolated experimental data, and numerical computations. In the range of interest there is as yet no generally accepted objectionless description that represents an optimum and "cost effective" mix of analysis and algorithmic modeling.

Among the criticisms that have been leveled at current approaches to high power microwave-atmosphere interaction problems are:

(a) The space-time dependent absorption and reflection electromagnetics of an incident microwave pulse are not treated accurately; reflections as well as transverse to longitudinal mode coupling are either neglected completely or are taken into account by ad hoc procedures of questionable accuracy.

(b) The non-equilibrium nature of the particle kinetics is not adequately reflected in fluid descriptions of the various particle species.

(c) In the energy range of interest collision parameters representative of the atmospheric chemistry and descriptive of momentum and energy transfer, ionization, attachment, recombination, heat conductivity, etc. are derived with questionable accuracy by extrapolation from a low energy base. There is also a question as to the optimum number of chemical reactions necessary to achieve engineering accuracy.

Although the ultimate microwave pulses of interest must have peak powers in the vicinity of static breakdown (10^6 watts/cm^2 at ground level) and microsecond or longer pulse lengths, only two facilities within the United States are presently under construction to generate such pulses, one partially completed at the Naval Research Laboratory and one planned at General Dynamics. The tests scheduled for these facilities tend to be global in nature, i.e., they demonstrate overall effects, but do not follow individual interaction processes that elucidate the basic phenomena. Therefore, in view of the substantial time delay before gigawatt facilities become operational and the global character of the scheduled very high power tests, experiments at lower peak power levels will be highly desirable both to test out the evolving theory and to provide significant program direction.

Although an impressive amount of research, both analytical and simulated, has been carried out on space-time dependent descriptions of wave-matter interactions by workers in this country and by the Russians, there is comparatively little that has been done in the microwave frequency range.

Over the past few years our group at the Polytechnic has investigated high-power, short-pulse microwave propagation in the atmosphere with support from JSEP, NAV AIR and NRL. One of our contributions was the first detailed space-time analysis and computer display of high-power microwave pulse propagation in the presence of atmospheric breakdown. This point was noted during a meeting of the DoD special committee on microwave technology, on which one of our group (Professor Marcuvitz) was a member. The experience gained from these involvements in microwave absorption processes in the atmosphere has underscored the need for addressing the criticisms indicated above.

Apropos of item (a) above, we have considered the application of quasiparticle techniques to absorption and reflection of pulse modulated microwave carriers by an inhomogeneous dielectric medium; there are well known analytical difficulties in the description of the reflected pulse because of the double frequency coupling terms arising in the inhomogeneous region. We have investigated a number of different techniques for averaging out the complicated high frequency space-time dependent terms so as to retain the important low frequency terms that determine energy-momentum transport. In particular, we have calculated overall energy transport over a large parameter range by concentrating on far field reflection and transmission coefficients. The accuracy of these calculations has been checked by comparison with numerical computations, at least in the range wherein computer runs are economical in terms of memory and execution time requirements. More recently we have explored analytically and numerically energy conversion processes associated with the interaction of a high power modulated microwave pulse and an atmosphere ionized by the pulse. The basic problem is that of a pulsed beam of finite spatial width and temporal extent incident

obliquely on an atmosphere at power levels such that the plasma formed is nonlinear, turbulent, and space-time variable. As noted above, analysis of this problem is dependent on the derivation of suitable equations descriptive of the electromagnetic and particle dynamics, and capable of being manipulated into a form for efficient numerical analysis. To obtain computational efficiency we have employed quasiparticle methods to describe both the microwave fields and the collective processes excited in the plasma formed by the pulse; these methods average out fast space-time oscillatory behavior and retain only the relatively slow space-time variability of engineering interest. As noted in Sec. B, a number of modular problems have been considered to obtain accuracy checks on our computational algorithms.

In connection with the item (b) above, we have investigated differences between results of kinetic and fluid model descriptions of an electron plasma under different types of excitation. It is known that collisionless kinetic models of particle systems can be approximated by unviscid nonlinear fluid models that under certain types of excitation lead to shock phenomena not present in the original kinetic model. This discrepancy, which appears to stem from the non-equilibrium nature of the particle kinetics, can be eliminated at least in certain ranges by the introduction of anomalous heat conductivity collision terms into the otherwise collisionless fluid model. However, the presence of sufficient particle-particle collisions of sufficiently large pulsed electric field excitation seems to obscure the above discrepancy. We have explored numerically parameter ranges in which the above discrepancy is or is not important.

In the predecessor work unit, entitled "Wave-Matter Interactions," general approaches were developed for "collision" (interaction) terms required in the description of particle-particle, particle-wave and wave-wave interaction processes which arise at those high power levels at which ionization and related phenomena take place during the propagation of an electromagnetic wave through various media. A common underlying analytical theme has been to obtain a kinetic basis for models of "collisional" interactions. Although the research reported there has been applied mostly at laser frequencies, the general approach in terms of a quasiparticle treatment of wave propagation through nonlinear media[1,2,3] is also applicable at microwave frequencies.

We have applied this approach to specific electromagnetic wave-media interactions, such as ionospheric scintillation and turbulence,[4-8,26] laser generation and propagation,[9-12] interactions with metallic surface,[13-19,22] to high-power microwave propagation and breakdown,[20,21] and in the past period to a number of related applications.[23-27] Investigation of the basic interactions in different electromagnetic frequency regimes has been helpful in the development of both analytical treatments and computer models for comparison with experiment.

To validate the microwave propagation computer model, we have assembled a conventional S-band rectangular waveguide system at the Polytechnic. A section of waveguide between two mylar windows is connected to a vacuum pump so that gas composition and pressure can be varied as desired. Air breakdown and pulse shortening are observed when the pressure within this waveguide section is reduced from atmospheric to a region in the vicinity of the Paschen minimum. In

Fig. 2, for example, pulse shortening of a 50-kW, 400-ns, 3-GHz pulse is shown to take place when the pressure is between 2.1 and 20.2 torr.

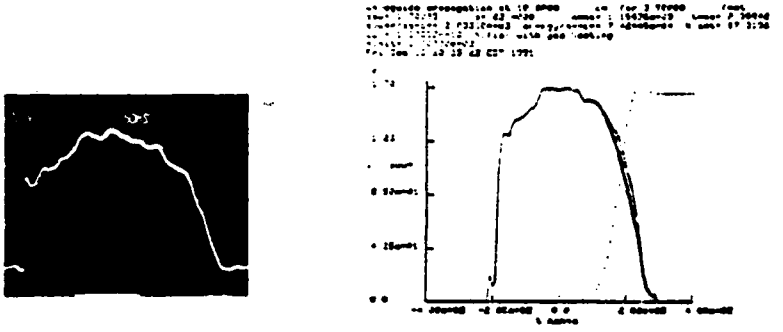
Each experimental picture in Fig. 2 is a double exposure. One exposure shows the pulse transmitted in the partially evacuated waveguide section when the waveguide pressure is reduced to the value specified under the picture. The second exposure shows the pulse transmitted when the waveguide contains atmospheric pressure. At 760 torr the field strength is too low to increase the electron density to a level where any modification of the pulse shape occurs. Also shown in Fig. 2 are the computer model results for propagation of a pulse at the specified pressures. The solid curves show the pulse shape incident at the teflon window as well as after propagating 3.5 feet in the waveguide section with reduced pressure. In the results shown in Fig. 2 the actual experimental pulse shape was digitized into the computer and used in the model.

Experimentally, pulse shortening is observed in Fig. 2 within a pressure region of 2.1 to 20.2 torr. Above 2 torr the agreement between experiment and computer model is good; however, below 2 torr the model results also indicate erosion of the pulse tail which is not observed experimentally. In earlier experimental-computer model comparisons using a shorter 80 ns pulse, good agreement was obtained over the entire 1 to 4 torr pressure region within which pulse shortening was observed by small modifications of several empirical coefficients.

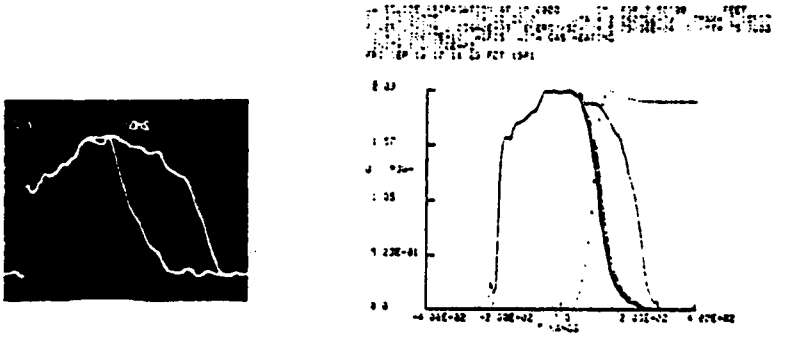
A sensitivity analysis of the empirical coefficients has begun. The initial results indicate that the momentum transfer coefficient is the most sensitive. A 20% change in this coefficient produces twice the effect on pulse shape that a similar change in the energy to momentum transfer ratio coefficient or the average energy coefficient does. The ionization coefficient is less sensitive. A 20% change produces only half the effect on onset of pulse shortening.

Thus far, simple functional forms (linear, exponential, etc.) have been used for the empirical coefficients in the model. We have begun to examine the effect of utilizing closer fits of the empirical coefficients to the underlying data base. A momentum transfer coefficient modified to more closely fit the data base decreased pulse shortening at all pressures. A modified energy/momentum transfer coefficient increased pulse shortening at all pressures. A modified average energy coefficient had little effect at the low pressure end but decreased the pulse shortening at the high pressure end. These modifications have not been incorporated into the model thus far. We may anticipate further changes in the empirical coefficients following an up-dating of the underlying data base.

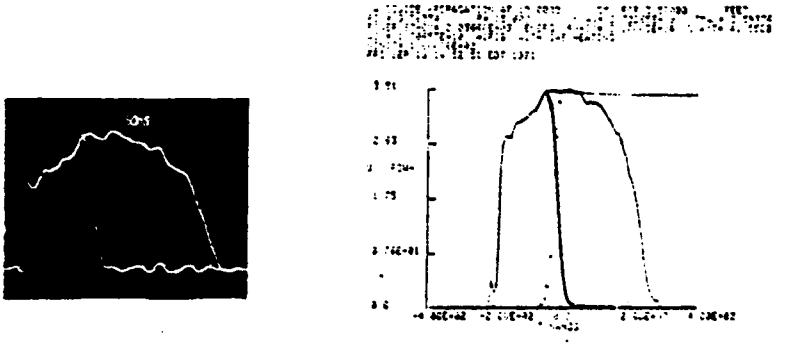
Development of the computer model for the propagation of a high-power microwave pulse through the atmosphere has continued. Initial interest was in shorter nanosecond pulses with peak powers substantially above that required for dc or static breakdown for applications involving sneak-through propagation. More recently, interest has broadened to include longer microsecond pulses with peak powers only slightly above static breakdown values. Such longer pulses may be useful in producing a dump of energy into a breakdown plasma at a substantial range from the source. Gas heating and hydrodynamic effects will play important



PRESSURE 22.4 TORR

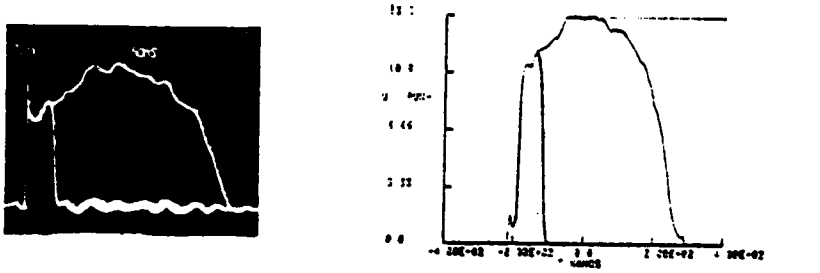


PRESSURE 20.2 TORR

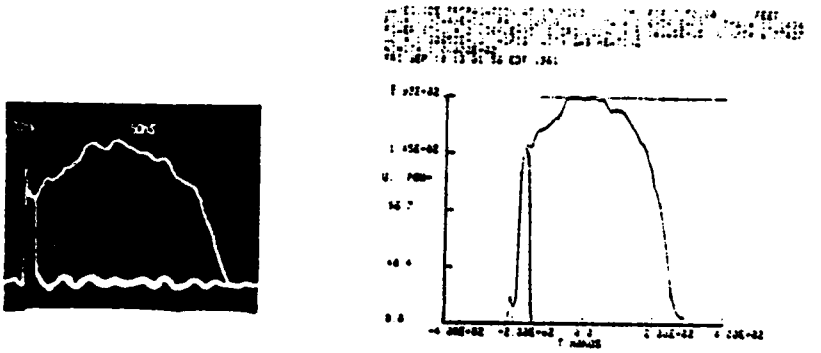


PRESSURE 15.6 TORR

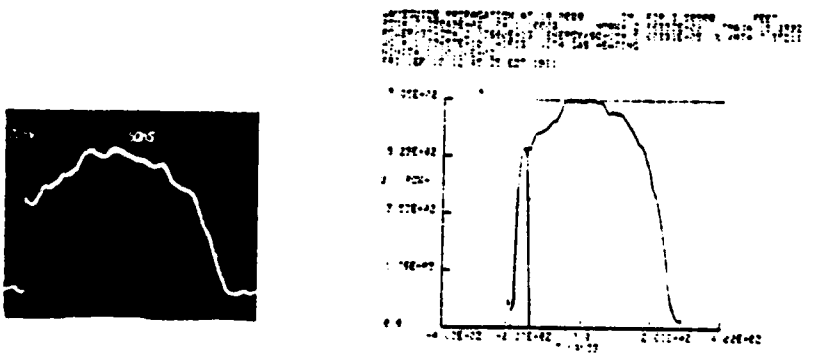
Fig. 2a Comparison of experiment with computer model results for propagation of a 3-GHz, 50-kW 400-ns microwave pulse in a partially-evacuated waveguide.



PRESSURE 8.0 TORR



PRESSURE 2.1 TORR



PRESSURE 1.1 TORR

Fig. 2b Comparison of experiment with computer model results for propagation of a 3-GHz, 50-kW 400-ns microwave pulse in a partially-evacuated waveguide.

roles. Provision has, therefore, been made in the computer model for the temperature of the atmospheric gases and to allow for their heating by the propagating pulse.

Extension of our computer model to follow the propagation of longer pulses has been hampered by insufficient computer storage space and too slow speed of our 16-bit MRI computer, a DEC PDP 11/60. As described in Sec. 3, the Summary of Recent Progress, we have been correcting this situation by adding 32-bit word processors to our MRI Computational Facility; first a network of four Apollo computational nodes and more recently a Gould super minicomputer with a CSPI array processor. JSEP equipment funding was responsible for the Gould-CSPI addition which is expected to provide additional storage space as well as increased speed to be able to extend our computer model to longer pulses.

Two additional processes must be considered in completing the validation: (1) the initial electron density to be used in the computer model and any residual-charged or easily-ionizable species remaining from a previous pulse, and (2) heating of the waveguide window producing initiators of breakdown. Initial experimental data was taken at a pulse repetition rate of 357; in Fig. 2 the rate was reduced to 8 pps. The initial electron density used in the computer model results was 100 cm^{-3}. Substantial reductions (e.g., to 10, 1 or 0.1 initial electrons cm^{-3}) do produce significant changes in the shape of the transmitted pulse.

Previously, our 50 kW magnitron provided pulse shortening data in a narrow pressure region around the Paschen minimum, as indicated in Fig. 2. As described in Sec. 3, a new high-power modulator has been constructed and a new 1 MW magnetron with longer pulse capability has been acquired.

5. REFERENCES AND PUBLICATIONS

1. N. Marcuvitz, "Quasiparticle View of Beam Propagation in Nonlinear Media," Nonlinear Electromagnetics, ed. P. Uslenghi, Academic Press, p. 117-132 (1980).

2. N. Marcuvitz, "Nonlinear Wave Propagation," URSI International Symposium on Electromagnetic Waves, Munich, Germany (September 1980). Invited.

3. N. Marcuvitz, "Quasiparticle View of Wave Propagation," Proc. IEEE, Vol. 68, p. 1380 (November 1980). Invited.

4. S.R. Barone, "Nonlinear Theory of Type II Irregularities in the Equatorial Electrojet," Phys. of Fluids, Vol. 23, p. 491 (March 1980).

5. S.R. Barone, "Renormalization of Maxwell's Equations for Turbulent Plasma," Phys. Rev. A, Vol. 21, p. 1725 (May 1980).

6. D.M. Wu, "Ionospheric Scintillation," Ph.D. Thesis, Polytechnic Institute of New York (August 1980); also POLY-MRI Report No. 80-1410.

7. S.R. Barone, "Angular Momentum of Particles in Geomagnetic Field," Phys. Letters, Vol. 80a (December 4, 1980).

8. N. Marcuvitz and D. Wu, "Ionospheric Scintillations," URSI Boulder Meeting (January 1981).

9. M. Newstein and F. Mattar, "Adaptive Stretching and Rezoning as Effective Computational Techniques for Two-Level Paraxial Maxwell-Bloch Simulation," Computer-Phys. Commun., Vol. 20, pp. 139-163 (1980).

10. G.M. Kull, "Analysis of the Copper Vapor Laser and Development of a General Laser Plasma Model," Ph.D. Thesis, Polytechnic Institute of New York (May 1980).

11. W.T. Walter, N. Solimene and G.M. Kull, "Computer Modeling to Direct Copper Vapor Laser Development," Proc. of International Conference on LASERS '80 (C.B. Collins, editor); STS Press: McLean, Virginia, pp. 148-158 (1981).

12. W.T. Walter, "Stepwise Excitation - A Limiting Process in the Copper Vapor Laser," Proc. of International Conference on LASERS '81 (C.B. Collins, editor); S.T.S. Press, McLean, Virginia, pp. 853-865 (1982).

13. W.T. Walter, "Reflectance Changes of Metals During Laser Irradiation," Proc. of SPIE, Vol. 198, ed. J.R. Ready and C.B. Shaw, Jr., pp. 109-117 (1980).

14. N. Solimene and M. Newstein, "Laser-Metal Interaction in Vacuum," XI International Quantum Electronics Conference, Boston (June 1980).

15. T.H. Kim, "High-Energy Pulsed Laser Interaction with Metallic Surfaces," Ph.D. Thesis, Polytechnic Institute of New York (June 1980).

16. M. Newstein and N. Solimene, "Laser Metal Interactions in Vacuum," to appear in IEEE J. of Quantum Electronics, Vol. QE-17, pp. 2085-2091 (October 1981).

17. M. Newstein, "Laser Interaction with Metals," Proc. of International Conference on LASERS '80 (C.B. Collins, editor); STS Press: McLean, Virginia, pp. 71-80 (1981). Invited Paper.

18. W.T. Walter, N. Solimene, K. Park, T.H. Kim and K. Mukherjee, "Optical Properties of Metal Surfaces During Laser Irradiation," in Lasers in Metallurgy (K. Mukherjee and J. Mazumder, editors) TMS-AIME Conference Series, pp. 179-194 (1982). Invited Paper.

19. K. Mukherjee, T.H. Kim and W.T. Walter, "Shock Deformation and Microstructural Effects Associated with Pulse Laser-Induced Damage in Metals," in Lasers in Metallurgy (K. Mukherjee and J. Mazumder, editors) TMS-AIME Conference Series, pp. 137-150 (1982). Invited Paper.

20. N. Marcuvitz, N. Sclimene and W.T. Walter, "High-Power Microwave Propagation," POLY-MRI Report No. 1413-80 (December 1980).

21. W.M. Bollen, W.M. Black, R.K. Parker, W.T. Walter and R.T. Tobin, "High Power Microwave Breakdown of Gases," Plasma Physics Division of American Physical Society, San Diego (November 1980).

22. W.T. Walter, N. Solimene, K. Park, T.H. Kim and K. Mukherjee, "Optical Properties of Metal Surfaces during Laser Irradiation," Proc. of International Conference on LASERS '81 (C.B. Collins, editor); STS Press, McLean, Virginia, pp. 510-524 (1982).

23. G. Kopcsay, "Propagation of Pulses in Nonlinear and Turbulent Plasma," Ph.D. thesis in preparation.

24. S.Y. Lee and N. Marcuvitz, "Beam Reflection from Lossy Dielectric Layers," Journ. of Optical Society, Vol. 73, No. 12 (December 1983).

25. S.Y. Lee and N. Marcuvitz, "Quasiparticle Description of Pulse Propagation in a Lossy Dispersive Medium," Trans. on Ant. and Prop. (April 1984).

26. D. Wu and N. Marcuvitz, "Ionospheric Scintillations," Radio Science (October 1984).

27. N. Marcuvitz, "Quasiparticle Method for Propagation in Inhomogeneous Media," Proc. of the 1983 URSI International Symposium on Electromagnetic Theory, pp. 33-36, Santiago de Compostela, Spain (August 23-26, 1983).

6. DoD AND OTHER INTERACTIONS

About three years ago, Professor Marcuvitz was contacted by R. Wasneski of NAV AIR about our willingness to investigate the possible realization of atmospheric channels created by and capable of propagating very-high-power microwaves for distances of a few kilometers. Such a possibility had been reported in a Russian publication. A quick NAV AIR sponsored study indicated that channel creation was not feasible with currently available microwave sources. Other possibilities, however, for the transmission of high-power microwave pulses were suggested during the study.

The background for our quick response stemmed from some of the basic nonlinear and turbulent wave propagation studies that originated from our JSEP research. More specifically, our JSEP program on "Wave-Matter Interaction" provided basic technical contributions for our recent NRL-sponsored program on "High Power Microwave Pulse Propagation," which was both theoretical and experimental.

As a related consequence of this program, Professor Marcuvitz was invited to become a member of a special DoD committee to undertake the development of a preliminary plan for a U.S. High-Power Microwave Technology (HPMT) program with potential applications to DoD missions.

SECTION II
SOLID STATE ELECTRONICS

SECTION II: SOLID STATE ELECTRONICS

A. ENHANCEMENT OF INELASTIC OPTICAL PROCESSES ON SMALL AND ULTRA-SMALL SOLID STRUCTURES

Professors S. Arnold, H. J. Juretschke, A.A. Oliner and P.S. Riseborough

Unit SS3-1

1. OBJECTIVE(S)

The initially-astonishing phenomenon of surface-enhanced Raman scattering (SERS) was discovered about three years ago, when it was found that the Raman scattering from selected active molecules was increased by seven or so orders of magnitude when the molecules were placed on a rough silver surface. A large number of experimental, and some theoretical, investigations followed; although there is still disagreement with respect to the physics of the effect, most research s believe it is due primarily to surface plasmon resonances of microsc metal particles in the optical range, even though full understanding s obviously not been achieved. Very recently, it is being recogn that similar spectacular enhancements can be achieved for a large v ety of other inelastic optical phenomena, such as fluorescence, ph emission, second harmonic generation, photocatalysis, photopho forces, and so on. In addition, specialized geometries, such as p iodic arrays, are being proposed to deliberately sharpen the resonances and improve the efficiency of the process. We believe that we are on the threshold of an exciting new field, with enormous potential for both the production or enhancement of new inelastic optical effects and the ability to exploit in various ways the resonances which produce these enhancements.

Our broad objective is to explore, understand, and optimize these enhancements of inelastic optical processes on small and ultrasmall structures through the use of well-defined geometrical configurations where experiment and theory allow the identification of the contributions to the local field buildups, among other mechanisms.

2. APPROACH

We are proposing a combined experimental and theoretical investigation of the fields and enhancements in a number of optical processes at the surfaces of small spherical particles, of plane surfaces with controlled superimposed roughness, and of non-spheroidal structures, and of their arrays. The experimental probes will include fluorescence, second harmonic generation and photoemission; we do not plan to do Raman scattering.

For the measurements on small spherical particles, we plan to use a novel facility that can levitate a single small spherical particle, and which has a sensitivity that can detect the departure of a single electron and a stability whereby the particle's charge value can remain constant within an electron for as long as a day. The experiments involving controlled roughness will take advantage of the thin film facilities and expertise that apply to Work Unit SS3-2. The theoretical modeling will concentrate on useful and clear formulations of electro-

magnetic field resonances in such geometries and their extensions to array and other environments, and on quantum-mechanical contributions to such local fields.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

Major advances have been made in the characterization of the interaction of electromagnetic energy with single submicroscopic spherical particles, and these studies have led to several new and basic results. First, we have constructed a simple physical description of the anisotropy of energy absorption within the particle, and we show that it can, in fact, produce a reversal in the direction of the photophoretic response of such particles. Incidentally, this description indicates that the commonly assumed validity of the Rayleigh scattering description of very small particles is not always applicable. Under the appropriate resonance conditions, the total force due to radiometric and photon pressure effects can also reverse, and we predict that this effect should become very large for semiconductor particles.

Second, we have developed a new method for studying the absorption of energy in single small particles. This method relies on the fact that such energy absorption produces a small change in temperature which, in turn, results in a change of particle size. Changes of size as small as 10^{-2} Å have been detected by us, using the shift with size of the sharp surface resonances of a small sphere in a non-absorbing region of the optical spectrum. Consequently, this new method of Structure Resonance Modulation Spectroscopy allows, for the first time, a direct study of energy absorption in single particles of very small mass, of the order of picograms.

Third, in connection with related experiments on photophoresis, the theory of the particle size dependence of the photophoretic force has been corrected and extended beyond the continuum approximation of hydrodynamics, that is commonly used to treat the interaction of a small particle with the surrounding gas. Even for particles as large as four times the gas mean free path we have found significant corrections to the interaction. Such corrections should play a role in all cases where these two dimensions are comparable. In particular, they have enabled us to obtain from the experimental data of the photophoretic force an absolute measure of the effect of electromagnetic radiation on the particle. This now agrees fully with the prediction of that effect based on a first principles calculation.

In the project to observe the onset of nonlinear optical properties on nearly perfect metal surfaces, a delay was encountered because the current vaccum pumping system proved to be incapable of handling the gas load during sample deposition. Consequently the system is not sufficiently clean to obtain surfaces of the requisite degree of perfection. This problem is being corrected, within the rather stringent criteria that the whole system must be portable to have access to various light sources.

A self-consistent quantum mechanical treatment of the electromagnetic interaction with a very thin metal slab, undertaken in order to probe for the onset of non-classical effects, has been carried out. It shows that there are resonances in the reflectances not present in any semi-classical treatment, but also that these resonances are extremely sensitive to broadening by the presence of a finite electron mean free path. The same type of calculation has been pursued to treat a semi-one-dimensional conducting system, where a similar sensitivity to the degree of electron scattering was encountered.

4. STATE OF THE ART AND PROGRESS DETAILS

A. General Background

The recent finding that the intensity of Raman scattering in certain molecules is enhanced by many orders of magnitude (SERS) when the molecule sits close to a solid surface having particular structural characteristics, has let loose a flood of investigations in order to probe a great variety of different experimental aspects of this phenomenon, and to build theoretical models for explaining these spectacular observations. Without going into detail in either experiment or theory, two general points are clearly emerging from this work. First of all, this effect is only one of many possible optical interactions taking place at surfaces that will show considerable enhancement. Secondly, by effectively extending the range of the strong coupling of electromagnetic waves with structures to a domain where the physical size of the structure becomes small compared to the wavelength, such interactions are bound to lead to novel practical applications involving various aspects of electromagnetic radiation. With the perspective of these two points in mind, we propose here to carry out a new study in this field. It builds on our considerable backgrounds in both electromagnetic wave interactions and in surface properties gained in part through past and current JSEP programs, as well as on a novel and highly sensitive measurement technique for probing optical effects at well-defined surfaces.

The enhancements (of 10^6 or so) observed in SERS are widely documented, though often only in sketchy form and with only rather qualitative background information, but even review articles[1,2] can hardly keep pace with the new findings. The probing of enhancements in other optical interactions is only beginning. To site a few significant examples, there is the 100-fold enhancement of photoemission yield from less than 50Å diameter silver particles observed in a sol,[3] the second harmonic generation, of enhancement 10^4, from metal surfaces chemically roughened to a depth of 500 Å,[4] and the observation of strong luminescence of dye molecules on metal islands, in contrast to the quenching of this radiation on flat metal surfaces.[5]

Many of these results are not fully understood, but it is usually agreed that one or more of the following factors must play a role. Firstly, the proper local electromagnetic fields consistent with the nearby boundary conditions must be used for both the exciting and scattered radiation.[6,7] This is particularly important when either or both fields have frequencies that can excite electromagnetic modes in the solid related to geometrical resonances. Secondly, both fields may be subject to resonances in the dielectric response of the surface,

because of coupling to other excitations in the solid, such as extended surface plasmons, made possible by the surface structure.[8] Thirdly, the microscopic response of the surface, i.e., the form of the quantum-mechanically self-consistent electromagnetic fields within and close to the surface resulting from spatially-varying and non-local dielectric properties of the surface region, can be a sensitive function of the surface structure.[9] Finally, there are arguments that some enhancement is connected with special active sites of atomic dimensions.[10]

Understanding the relative role of these various contributions is essential if one wants to assess the general potential for enhancing inelastic optical phenomena by the use of structural surface features. We propose a study that, on the one hand, will build on canonical surface configurations to separate and identify various contributions to structural enhancement and, on the other, will develop and apply general criteria for utilizing the method for investigating otherwise very weak higher-order optical interaction effects. The work will be both experimental and theoretical.

The two surface configurations that form the starting point of the experimental program rely on unique expertise available in our laboratory. The surfaces are a= single spheres of dimensions comparable to and smaller in size than the wavelength of light, and b= nearly perfect, flat single crystal metallic surfaces. Professor Arnold, who has recently joined the Physics department and is currently an Alfred P. Sloan fellow, has perfected novel techniques for the stable suspension of single very small particles in a configuration that readily lends itself to investigating optical effects on their surfaces.[11] The extraordinary sensitivity of his instrumentation expressed in terms of current, is of the order of one electron per day, and with it he has already demonstrated the detection of striking effects on local optical fields with changes in particle size as small as a few angstroms.[12] These spheres provide a well-defined and theoretically-tractable geometry which allows unambiguous and quantitative evaluation of both classical and quantum mechanical influences on the local optical field. This is in contrast to the usual experiments on rough surfaces where an undefined fraction of the surface is active, and under only grossly-defined geometrical conditions.

Professor Juretschke has specialized in the preparation of metallic surfaces that are flat and perfect to the extent of being practically inert to interaction with an environment. Such surfaces will serve as a starting point for studying the effects of the controlled and deliberate introduction of structural defects, ranging from point sites to surface dislocations, to hillocks and periodic modulations. Again, the initial geometry is well defined, and the roughening can be controlled.

The theoretical effort is expected to be in strong support of the experimental program. It must concentrate on both the macroscopic and microscopic properties of any such structures, and on the local fields that are in consonance with these properties. It has an equally natural strength in this laboratory. Professor Oliner brings substantial insight and elegant mathematical techniques to the understanding of local electromagnetic fields in complex geometries. This will complement Professor Riseborough's wide background in many-body theory and transport phenomena -- he recently came to the Physics department from the

University of California at La Jolla, San Diego -- for studying their microscopic dielectric response, taking quantum-mechanical effects into account.

As already mentioned, this proposal has two central themes.

(a) The unambiguous determination of the contribution of local field buildups to the observed enhancements, and the theoretical modeling of some other contributions,

(b) The development of general criteria for optimal enhancements of weak higher-order optical phenomena.

Within this program, the work will have four main thrusts:

1) Resonances and resonant optical effects at the surface of single nonmetallic particles (Arnold, Oliner, Riseborough)

2) Optical interactions at surfaces with controlled roughness (Juretschke, Arnold)

3) Electromagnetic theory of plasmon resonances on small non-spheroidal particles, including the effect of array and other environments (Oliner, Juretschke, Arnold)

4) Quantum-mechanical aspects of local electromagnetic fields close to metal surfaces and small structures (Riseborough, Juretschke)

This division is to some extent purely formal, and does not at all imply separate efforts by the different members of the work unit. As specified above, it is, in fact, expected that the main work within any subdivision will involve a very close collaboration by at least two of the four investigators, with the pairing shifting for the different subunits. Other interactions will, of course, develop naturally as the work proceeds.

The four main areas of research will be discussed in general way below, and the detailed program proposed in each area is presented in a subsequent section. Because of the unique nature and novelty of the measurement facility for the properties of single particles, this technique is discussed in a section of its own.

B. Resonances and Resonant Optical Effects at the Surface of Single Nonmetallic Particles

(1) Introduction

A single sphere is undoubtedly the simplest small structure for which the interaction with an electromagnetic plane wave can be fully specified. It is therefore also an ideal geometry in which to search for contributions to optical interactions on its surface that go beyond those arising from a purely electromagnetic field buildup. Before reaching that stage, however, a number of both experimental and theoretical questions must be clarified. For example, from the experimental side it has to be established that the effects of molecules covering only a fraction of the surface of a very small particle are, in fact, detectable

with any kind of sensitivity. Secondly, one also requires to have good first hand information about the optical properties of the bare particle itself, in order to be able to interpret the results in a meaningful way. On the theoretical side, it is equally important to have quasi-intuitive guidelines for understanding the change in the response of the sphere as the various parameters of the problem of the interaction with electromagnetic waves are changed, above and beyond obtaining the complete but usually not very transparent, and entirely computer-dependent, formula solutions. We will outline here some answers to these questions, show that an approach along these lines will, in fact, be feasible, and thereby lay the groundwork for more detailed investigations. It will become clear, though, that these methods alone open up new ways of determining the properties of small particles that remain to be exploited in their own right. We will concentrate on discussing nonmetallic particles, although many of the points also apply to semiconductors or metals.

While the straightforward detection of molecules on the surfaces of small dielectric particles is limited by their extremely small number, we expect that, just as in the case of Raman enhancement on metals, local resonances can improve the response appreciably. For example, in the case of dielectrics, enhancements can exist due to electromagentic surface wave resonances. Recently, Chang et al[13] have detected spikes in the external excitation wavelength dependence of emitted fluorescence from dye-coated optical fibers that correlate well with the surface wave resonant modes of these structures. Similar enhancements due to resonances are predicted for the emissivity spectrum of carbon-coated alumina spheres. According to Pluchino[14] a 10Å thick outer layer of carbon should increase the emissivity of a 6 micron alumina sphere by three orders of magnitude. Pluchino feels that surface waves are responsible for this enhancement; however, there are no substantiating experiments for his prediction or for the proposed mechanism.

A direct method for verifying this would not be easy. Unfortunately it is difficult to quantify fluorescence experiments, because of the sparse area in which fluorescence photons fall, and because the precise way in which an excited molecular adsorbate interacts with a curved dielectric surface is unknown. We propose that another and a better way to arrive at a quantitative understanding of such effects is through a thermal spectroscopy, since the partition of thermal energy between a particle and the outside environment is easily analyzed.

In addition to providing the proper tool for such thermal spectroscopy, the general technique making it possible will, in fact, be much more versatile, and also lend itself to probing a number of other small particle properties that are of interest to this proposal.

In light of the above, our objectives are two-fold:

(a) To establish a new experimental methodology for quantifying the interaction of electromagnetic waves with small particles, and with molecules on the surfaces of these particles.

(b) To use our traditional strength in electromagnetic theory at the Polytechnic in order to evolve a physical understanding of the numbers arrived at in (a).

(2) Experimental Methodology

In order to understand the manner in which dielectric spheres adsorb radiation we have established a unique probe. This probe obtains its information from the radiometric force on an illuminated particle in a gaseous atmosphere. This radiometric force, known as the photophoretic force, was discovered in 1917 by Ehrenhaft.[15] At a pressure for which the molecular mean-free-path is smaller than the particle size, the photophoretic force F_{ph} is given by

$$\vec{F}_{ph} = - CRJ\vec{I} \tag{1}$$

where C is the quantity which depends on well-known thermal and hydrodynamic properties of the particle and the gas, R is the radius of the particle, $\vec{I}$ is the incident light intensity, and J is a measure of the anisotropy of the internal heat sources. In particular[16]

$$J = Q_{abs} \; \frac{3}{8} \frac{\int_0^1 dx' x'^3 \int_0^\pi d\theta \, \sin 2\theta \, B(x',\theta)}{\int_0^1 dx' x'^2 \int_0^\pi d\theta \, \sin\theta \, B(x',\theta)} \tag{2}$$

where Q_{abs} is the efficiency for absorption, and $B(x',\theta)$ is the square of the electric field density at coordinate (x',θ) within the sphere (x' is a reduced radius, i.e., $x' = r/R$). The coordinate system for using Eq. (1) is shown in Figure 1. As one can see, the sign of J is controlled by the expression in the brackets. If the internal heat sources, that are proportional to B, are primarily on the side away from the excitation, J will be positive. In the opposite case, J will be negative.

For a small quantity of surface adsorbate, the angular distribution of sources $B(1,\theta)$ is expected to be the same as in the case without adsorbates so that a measurement of J provides a measurement of Q_{abs}. Inasmuch as emissivity and adsorption efficiency are equal through Kirchhoff's law, a measurement of J through the photophoretic force also provides a measurement of the emissivity E.

A second feature of this method is that one well-defined particle may be investigated at a time. This is accomplished by using a modified Millikan chamber as a precise force balance for the (slightly charged) particle.[11] Figure 2 shows the basic scheme of the balance. A more elaborate description of it will be given in the section titled Novel Measurement Facility for Single Particles.

The heart of the set-up is a modified Millikan capacitor (M). A nonuniform field is generated within this capacitor in order to create a potential well at the center of the plates. This in turn creates an unstable vertical potential, so that the particle (which is forced toward the center line) will attempt to move rapidly toward the top plate. The upward motion is overcome by readjusting the potential between the plates in accordance with the particle's position. The particle position is sensed by a position-sensitive detector (D) which responds to scat-

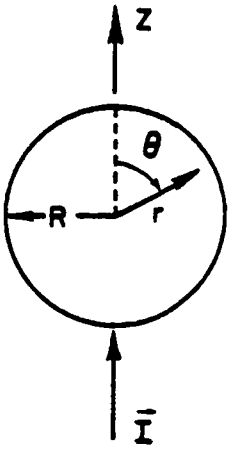


Fig. 1 Coordinate system for evaluating the integral J of Eq. (2) of the absorption of energy by a spherical particle of radius R illuminated from below with an intensity I.

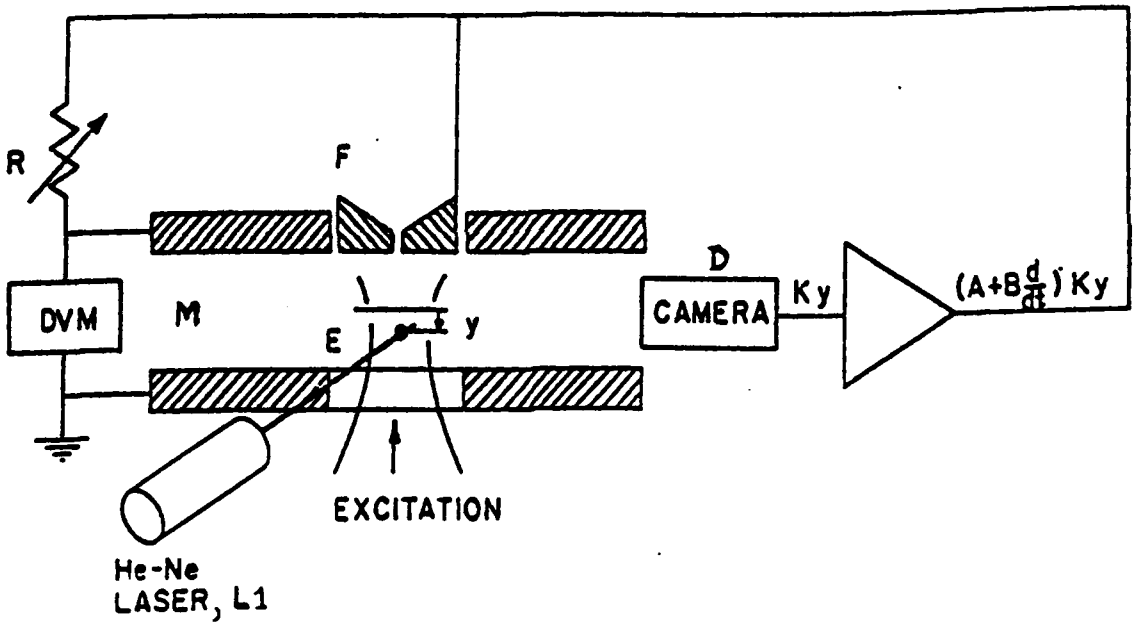


Fig. 2 Schematic of the automated Millikan chamber levitator for the stable suspension and study of single particles.

tered laser radiation (L1 He-Ne, 2mW). An overall servo-loop contains the levitated particle within tight constraints such that the particle changes its position by less than a diameter. Since the position is virtually constant, any photophoretic force F_p that is induced by radiation transmitted through the bottom plate (E) will automatically be balanced by a change in electric field between the plates: $F_p = -q\Delta E$. Here q is the particle charge. Since the weight alone was originally balanced by the initial field, $m\vec{g} = -q\vec{E}_0$, we obtain the relationship

$$\frac{F_p}{mg} = \frac{\Delta E}{E_0} \tag{3}$$

The photophoretic force is therefore measured absolutely from the particle weight and the fractional change in plate voltage. Along with the intensity $\vec{I}$ and the appropriate thermal and fluid properties, J is then easily found using Eq. (1).

Some preliminary results discussed below will show how the method can be applied. It should be noted that, while the charge q disappears in Eq. (3), it is obviously essential that q remain constant during the experiment. The conditions in the chamber can be arranged such that the particle charge stays stable within a single electron for as long as a day.

(3) Preliminary Results

A typical output of the computerized experimental analysis, directly giving the value of J, is shown in Fig. 3 for a slowly evaporating glycerol particle. The data was taken at an excitation wavelength of 10.63μ, with an intensity of 380 mW/cm^2. By keeping the wavelength λ constant and changing the particle size R as a function of time, we run through a continuous range of size parameter $\rho(= 2\pi R/\lambda)$, from $\rho = 15$ to $\rho = 3.5$. From $\rho = 15$ to 11.7 the force is "positive" (J is negative) and below $\rho = 11.7$ the force reverses and shows periodic structure corresponding to electromagnetic resonances. We will discuss in Section 9 why it is that the size parameter at which the force J changes sign gives the imaginary part of the refractive index, while the period of the ripples gives the corresponding real part. Hence the experiment yields directly the dielectric properties of a single small particle, a quantity that is normally very difficult to determine. In fact, there may soon be some interaction with the National Bureau of Standards on applications of this technique.

The data in Fig. 3 show that our experimental method can easily measure the quantity J. Since our primary interest is in determining the interaction of electromagnetic energy at a surface with adsorbed molecules, we must, in addition, verify that the photophoretic method can also detect a single monolayer or less.

For the purpose of obtaining such preliminary data a simple test was devised. A portion of a hydrosol of 7 micron polystyrene latex particles was deposited on a slide and dried. From the weight of this deposit the number of particles was determined. These particles were then placed in a solution of the dye Rhodamine 6G in water. The

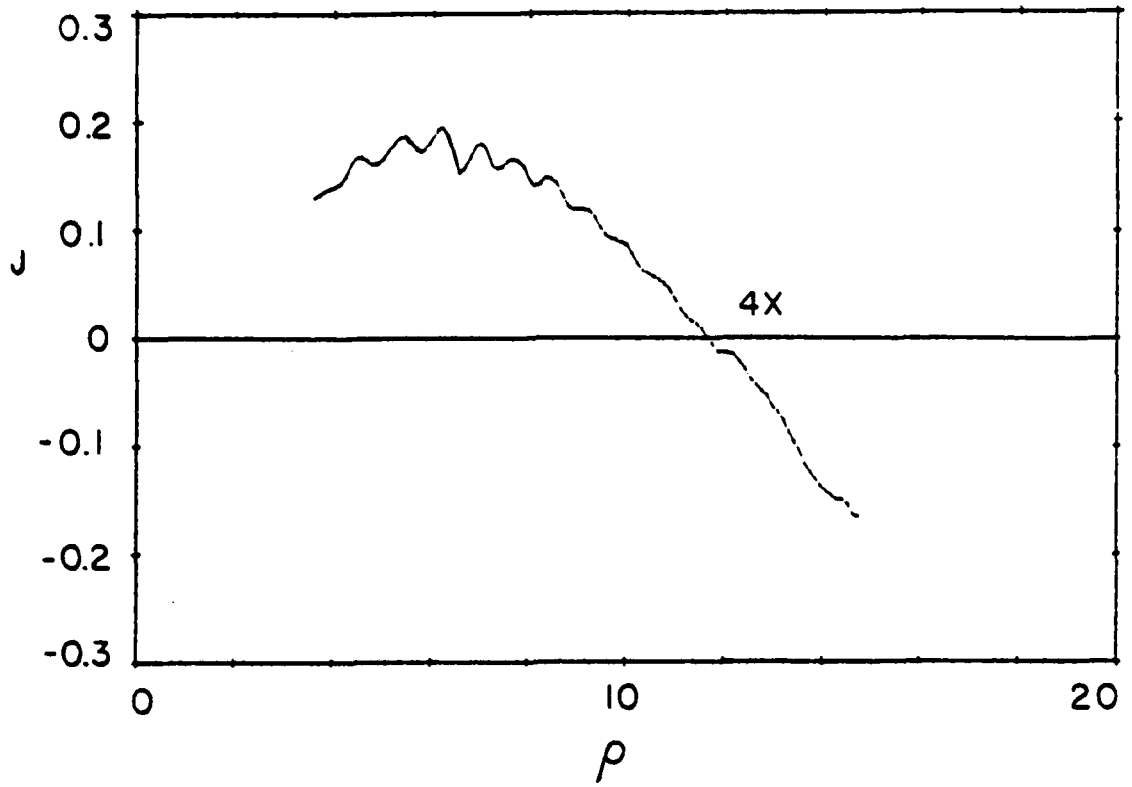


Fig. 3 The photophoretic energy absorption integral J of a slowly evaporating glycerol particle, radius R, as a function of the size parameter $\rho = 2\pi R/\lambda$.

concentration of this solution was such that if the dye were fully depleted in the adsorption process it would lead to no more than ~ 2 monolayers per particle. To assure lack of clusters the solution including particles was ultrasonicated for 5 minutes. After this the entire sol was centrifuged to separate the liquid solution. This supernatant was poured off into a cuvette and tested by absorption for loss of Rhodamine 6G. It was determined that less than 5% of the dye had been adsorbed. On average, each particle had taken up less than 10% of a surface monolayer.

A single Rhodamine 6G coated polystyrene latex particle was injected into our chamber and illuminated with Ar^+ laser radiation at 5145Å and 475mW/cm^2. At a pressure of 32 torr a photophoretic force-to-weight ratio $F_p/mg = -0.015$ was formed. The negative sign indicates that the force was opposite to the direction of the radiation ("negative" photophoresis). The force measured on several particles varied by as much as 30%, presumably as a result of slightly different particle sizes and surface conditions.

If Eq. (1) is used with the constant C as prescribed in Ref. 16, and with the known quantum efficiency of Rhodamine 6G, J is found to be 6×10^{-3}. This is a very large number when one considers that the approximate geometrical absorption efficiency Q_{abs} associated with 0.1 monolayer is 3×10^{-3}. In other words, the results suggest that the

asymmetric factor in brackets in Eq. (2) is approximately 2. Such a result is anomalous since the asymmetry factor in Eq. (2) cannot be larger than 1/2, and is very often a considerably smaller fraction.

One way out of this dilemma is to admit that geometrical thinking in arriving at Q_{abs} is incorrect. The geometrical picture asks, "how much light is absorbed in a single pass through the particle." This point of view ignores surface waves which, when in resonance, can stay in the proximity of a monolayer for a much greater length of time and therefore enable Q_{abs} to be enhanced. What is required for analyzing the present experiments is a complete Mie electromagnetic scattering calculation of the associated internal fields. Experimentally, surface waves should reveal themselves through resonant structure in the photophoretic excitation spectrum, either as a function of particle radius R, or of wavelength λ. In this preliminary experiment we did not determine whether or not resonance conditions were met.

On the basis of our preliminary experimental results we propose to

(i) Obtain systematic photophoretic excitation spectra on dye-coated dielectric microspheres in order to determine the mechanisms for the large value of Q_{abs} such as seen in our preliminary results.

(ii) Analyze these experimental results via a full description for the internal fields, in order to ascertain the contributions of the local electromagnetic field buildups. Such a process (Mie theory), if carried out purely formally, often leads to very little physical insight. Therefore we propose to construct a description which will identify the most important modal contributions to the surface fields and thereby hopefully obtain increased insight into the interaction, and expecially into the conditions for optimizing it.

(4) Simple Formulation of Electromagnetic Resonances for Spheres

Since the single particles under consideration here are of spherical shape, and since we are presuming that the phenomena involved here are predominantly, if not solely, electromagnetic in nature, these phenomena should be characterizable in terms of the classical Mie theory[17] for the scattering of an incident plane wave by a dielectric sphere of arbitrary size. However, the Mie theory phrasing of these problems appears enormously involved, since the plane wave is geometrically not compatible with a spherical coordinate system and must therefore be expanded into an infinite set of spherical waves, and numerical results are achieved only with the help of a complicated computer program. Various publications indeed proceed in this way in their theoretical treatments (e.g., refs. 14 and 18). What is lacking is a simple and transparent theoretical formulation that yields greater physical insight by excluding the unnecessary portions of the complete Mie formalism and retaining only the essential part or parts.

We shall seek to introduce and employ such simple and transparent formulations where possible by the use of spherical transmission line

theory, a rigorous technique in electromagnetics introduced in the 1950's at the Polytechnic and utilized in various contexts.[19] This approach should be particularly useful when we are interested in just the conditions for a specific resonance (and the field distribution corresponding to that resonance). Indeed, much of what we seek falls into that category. When we need the relative amplitudes of many different modes excited by a particular source, the formulation will necessarily become complicated no matter what phrasing is used.

Spherical transmission line theory is intended to parallel the common and widely-used uniform transmission line theory in its use of impedances, etc. An important similarity exists in that the modes can be divided into E modes and H modes in both representations. Important differences appear in the spherical case, however; for example, the characteristic impedance is a function of radius and is different looking towards or away from the origin. The directness of the spherical transmission line formulation may be illustrated by presenting a simple case here. Suppose we wish to write down the condition for resonance of a particular E mode; we then follow the standard transverse resonance requirement that states that

$$\overleftarrow{Y}_n(r) + \overrightarrow{Y}_n(r) = 0 \tag{4}$$

that is, the sum of the input admittances at radius r looking towards the origin and away from it must sum to zero. The form of these admittances depends on the radial character of the spherical particle; if the sphere is simply a bare homogeneous dielectric sphere then we choose r = R, if R is the radius of the sphere, and the admittances become the characteristic admittances which, for an E mode, are

$$\frac{\overleftarrow{Y}_n(R)}{\eta_1} = -i\,\frac{\hat{J}_n(k_1R)}{\hat{J}'_n(k_1R)} \tag{5}$$

looking towards the origin, and

$$\frac{\overrightarrow{Y}_n(R)}{\eta} = i\,\frac{\hat{H}^{(2)}_n(kR)}{\hat{H}'^{(2)}_n(kR)} \tag{6}$$

looking outward, away from the origin. The functions involved are the spherical Bessel functions, which are related to the ordinary Bessel functions of order (n + 1/2), and consist of combinations of sines and cosines. The prime signifies the derivative with respect to the argument,

$$\eta = \sqrt{\frac{\varepsilon_0}{\mu_0}} \quad , \quad \eta_1 = \eta\sqrt{\varepsilon_r} \quad , \quad k_1 = k\sqrt{\varepsilon_r} \tag{7}$$

where ε_r is the relative dielectric constant of the material comprising the sphere, and can be complex ($\varepsilon_r = \varepsilon' + i\varepsilon''$).

Substituting (5), (6) and (7) into (4), we obtain the rigorous resonance relation for any specific E mode as

$$\frac{\hat{H}_n^{(2)}(kR)}{\hat{H}_n'^{(2)}(kR)} = \sqrt{\varepsilon_r}\,\frac{\hat{J}_n(kR\sqrt{\varepsilon_r})}{\hat{J}_n'(kR\sqrt{\varepsilon_r})} \tag{8}$$

Condition (8) for the resonances of this class of spherical modes has been derived in a strikingly simple way, without all the clutter that customarily accompanies it in the usual treatments. We have no intention here of denigrating the usual treatments, which contain far more information and which aim to treat various other aspects, including the scattered far field, for which the complete treatment is required. We point out merely that when certain partial information is required, it is unnecessary to employ the complete analysis first and then extract the portion needed; one can derive the needed partial information simply and directly in many instances.

Suppose that the spherical particle is metallic in the optical range and is very small, and we seek the condition for resonance of the dipole mode. That case corresponds to the solution given in the literature[18] in an explanation of the large enhancements found in SERS. The mode in that case corresponds to n = 1, so that the spherical Bessel function ratios in (8) become

$$\frac{\hat{J}_1(s)}{\hat{J}_1'(s)} = \frac{\frac{\sin(s)}{s} - \cos(s)}{\sin(s)\left[1 - \frac{1}{s^2}\right] + \frac{\cos(s)}{s}} \tag{9}$$

where $x = kR\sqrt{\varepsilon_r}$, and

$$\frac{\hat{H}_1^{(2)}(v)}{\hat{H}_1'^{(1)}(v)} = \frac{-i\frac{u}{v} - u}{\frac{u}{v} + i\frac{u}{v^2} + i\frac{1}{u}} \tag{10}$$

where $v = kR$ and $u = \exp(ikR)$.

In the limit of very small R, (9) and (10) reduce to

$$\frac{1}{2}kR\sqrt{\varepsilon_r} \qquad \text{and} \qquad -kR$$

respectively, so that (8) yields

$$\varepsilon_r = -2 \tag{11}$$

in agreement with the literature.

The advantage here is not so much that one rather readily arrives at the well-known condition (11), which shows that this resonance on a metal particle can occur only when ε_r is negative, and also tells the wavelength of resonance according to the particular metal involved.

The use of (9) and (10) also permits us to find out quite directly how condition (11) changes as the particle size increases, and as the condition then becomes dependent on particle size.

The example given above is meant to be illustrative of how we hope to use spherical transmission line theory to yield needed partial information in a simple and direct way, so that numerical results may be obtained cheaply and physical insight may be more readily furnished.

(5) Previous Progress

The experimental program to determine the distribution of the local electromagnetic field around and within a microstructure has been developing in a number of directions.

The photophoretic determination of the anisotropy of the absorption of light within a spherical particle (the quantity J of Eq.(2), and its interpretation along the lines of Eq.(14)) has now been published[37]. In addition, these results recently received striking theoretical confirmation by A. Pluchino, who was able to locate both the oscillations and the crossover point in Fig. 3 on an absolute basis, using the complex dielectric constant of glycerol independently arrived at in Ref. 37. This confirmation assures the possibility of the complete determination of the complex dielectric constant of unknown microparticles by photophoretic measurements. From our point of view, this agreement shows that photophoresis also gives the information we are seeking concerning the field distribution within the particle.

As an interesting aside, the absolute magnitude of J as calculated is about 50% larger than the experimental value. This indicates that the actual thermal slip coefficient at the particle's surface is below the lower limit (0.75) set by classical kinetic theory, so that our data also contain information whose interpretation requires a more realistic understanding of the scattering of molecules at surfaces.

The extension of this method to probe local fields at or just above the particle's surface by using adsorbed molecules interacting with the electromagnetic field is in progress. The simplest configuration requires a known uniform coating of single molecules on the surface, and methods are being developed to deposit dye molecules without clustering and in controlled concentrations. The same theory that has confirmed the photophoretic interpretation of Fig. 3 also indicates that at resonance the internal field tends to be pushed toward the surface of the sphere. We therefore anticipate a huge enhancement in the photophoretic force for surface adsorbed molecules. Indeed, preliminary experiments indicate an enhancement of about two orders of magnitude. Such an enhancement should be accompanied by steep resonances in the force spectrum. We are currently searching for these resonances. The experiment involves the excitation of a surface coated micron-sized particle in the new quadrupole force balance, and was begun two months ago. The electromagnetic theory of this enhancement models the adsorbed molecules by a uniform dielectric coating, clearly only a crude approximation to the reality of individual surface adsorbed molecules. We will let the experimental results dictate the need for a more refined model.

We have worked out a method for determining directly the efficiency of absorption Q_a that enters as a factor into the expression J of Eq.(2). The method utilizes the natural electromagnetic resonances of a sphere. Excitation of any such a resonance, usually in the visible for a micron-sized particle, will produce a pronounced increase in backscattering. As an example, the predicted increase in backscattering is shown in Fig. 4 for a water droplet with a radius of 2.4980 μ.

This particular resonance, as seen in backscattering, extends over 10 Å in wavelength and can easily be swept through by a laser. Such a water droplet, which contains an involatile component, is in equilibrium with its vapor. An increase in temperature of the droplet by 10^{-3} °C brings its equilibrium size to 2.4977 μ, i.e., a decrease of 3 Å, which will result in the shifted backscattering spectrum also shown in Fig. 4. Hence a 10^{-3} °C change in temperature produces a significant change in backscattered intensity, which at the half-maximum point A is approximately 20%. These sharp resonances allow an extremely precise measurement of particle size and therefore of particle temperature. If the temperature increase is stimulated by light absorption, the shift in the backscattering spectrum provides a direct measurement of the amount of absorbed heat: Q_a is determined. The calculation in Fig. 4 assumes that no absorption takes place for the elastically scattered visible light. This new method for measuring absorption directly, what we have called Photothermal Modulation of Structure Resonances, thus senses the infrared absorption of a microstructure through the manner in which this absorbed energy modifies the visible elastically scattered light. Although our illustration applies to a liquid sphere changing its radius, the resonances can also be modified by alteration of the refractive index. For example, in the case of a 1 μ diameter semiconducting particle (CdS), the modulation in the position of the resonance is caused principally by the increase in refractive index with temperature. A paper on this method has been submitted for publication.[49]

The experimental realization of this new spectroscopy is under way and has already made considerable progress. It utilizes the new quadrupole electrostatic balance already discussed, which is operational, and performs satisfactorily for particles as small as 0.2 microns, with the feedback in balance optically controlled. Particles can be levitated very stably, and are spherical enough to show sharp structural resonances. Furthermore, neither the stability nor the resonances are strongly affected in resolution by steady-state temperature gradients of 10^{-3} °C due to heating of the sample. The predicted shift in backscattering with IR absorptions will be examined shortly.

Incidentally, the same method for independently determining Q_a will be needed for analyzing the optical response of particles with absorbing surface molecules.

In another advance, we have shown experimentally that small particles can be completely levitated without the need of an electrostatic balance in the properly arranged mode of a visible laser light.[50] As shown in Fig. 5, this can occur in partial doughnut modes, with the particle seeking the field minimum at the center. Depending on whether the particle is weakly or strongly absorbing, the interior tem-

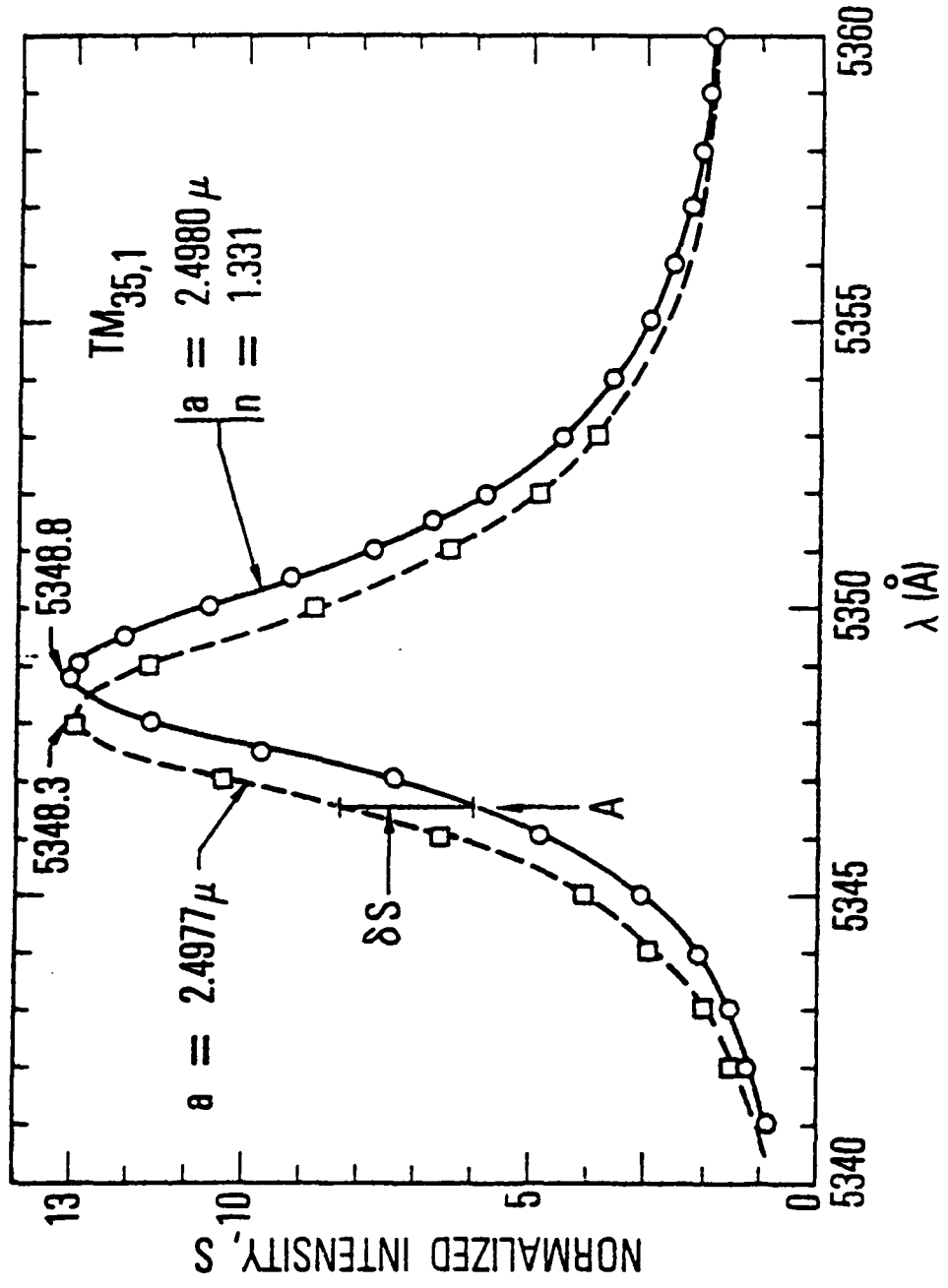


Fig. 4 Shift of backscattering of small particle with temperature.

perature gradient will oppose or be along the direction of the incident light, so that both forward and reverse levitation can occur. The significant features of the interactions are: 1) only very weak light

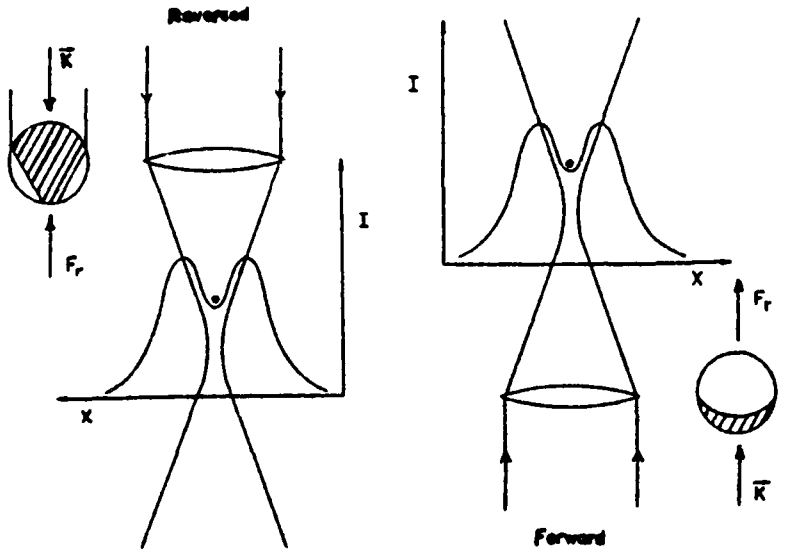


Fig. 5 Arrangement for laser levitation of small particles in a doughnut mode.

sources are needed for levitation, surely that, for example, ambient light levels in the atmosphere will suffice for such levitation of aerosols; 2) the stability of the lateral localization of very small particles is considerably enhanced and will therefore allow an even more precise determination of J. The experiments show that, although the particles are 40x the wavelength of the radiation, calculations based on geometrical optics fall far short of explaining the force for weakly absorbing particles (i.e., particles for which the product of the imaginary component of the refractive index times the particle size is much smaller than 1). Unfortunately, at these sizes (i.e., 40x the wavelength) full Mie calculations are too time-consuming at current computer speeds utilizing our new quadrupole levitator in which experiments may be performed at size parameters for which calculations are practical.

(6) Recent Progress

The recent work can be roughly ordered into three topics, which will be taken up in turn below.

a) Photophoresis

The understanding of our results on photophoresis, as shown in Fig. 3, and presented in a publication,[37] has been accelerated by new theoretical developments. Calculations published by

Pluchino,[47] based on Yalamov's theory of photophoresis,[16] reproduce the resonant features in Fig. 3, and yield a reasonable crossover of J as a function of size. But the actual force is less than that calculated, and this discrepancy increases to over a factor of two with decreasing particle size. Although Pluchino originally attributed this discrepancy to the use of Gaussian beams in our experiment, since the theory is worked out for plane wave excitation, this supposition proved to be incorrect. We have now found, in direct collaboration with Pluchino (at Aerospace Corp.) that the correct origin of the disparity lies in the breakdown of continuum hydrodynamic theory when applied to the combination of gas pressure and particle size in our experiments. Although the mean free path in the surrounding gas is never larger than one quarter of the particle radius, the measured force is less than one half of that calculated from continuum theory. To include modifications of continuum theory that remove this discrepancy, we have evolved a comprehensive new theory of photophoresis.[48] The overall result of this work leads to multiplying the right side of Eqn. (1) by a term $g(K_n)$ dependent on the Knudsen number K_n (ratio of gas mean free path to particle radius). The overall force then becomes

$$F_\rho = - cg(K_n)RJ\vec{I} \tag{12}$$

with

$$g(K_n) = \frac{1}{[1+3c_mK_n]\ [1+2c_tK_n]}$$

where c_m is related to momentum accommodation at the gas-particle interface (a number experimentally between 1.00 and 1.35), and c_t is the temperature jump coefficient (a number between 1.875 and 2.48). Fig. 6 shows the J derived from the photophoretic force using Eq. (12), and compares it with the theoretical value of J. The current agreement between theory and experiment is within 10%.

It is interesting to note that the best fit to the data requires a thermal slip coefficient of 1.25, which is in marked disagreement with the value 3/4 predicted by Maxwell on the basis of classical kinetic theory. Our data clearly contain information whose interpretation requires a more realistic understanding of the scattering of molecules at surfaces. An improved theoretical approach to describe such scattering is currently in progress.

b) Internal Field Anisotropy

It is clear from our data on photophoresis that well-developed resonant modes exist within any small microsphere. Many of these modes are of sufficiently high order that calculations of the anisotropic internal field distribution require a complicated polynomial expansion, and physical insight is difficult. The situation would be more understandable as we let the optical size become very small ($\rho << 1$) so that the theory need only retain the first term which can give rise to an anisotropy in the internal field.

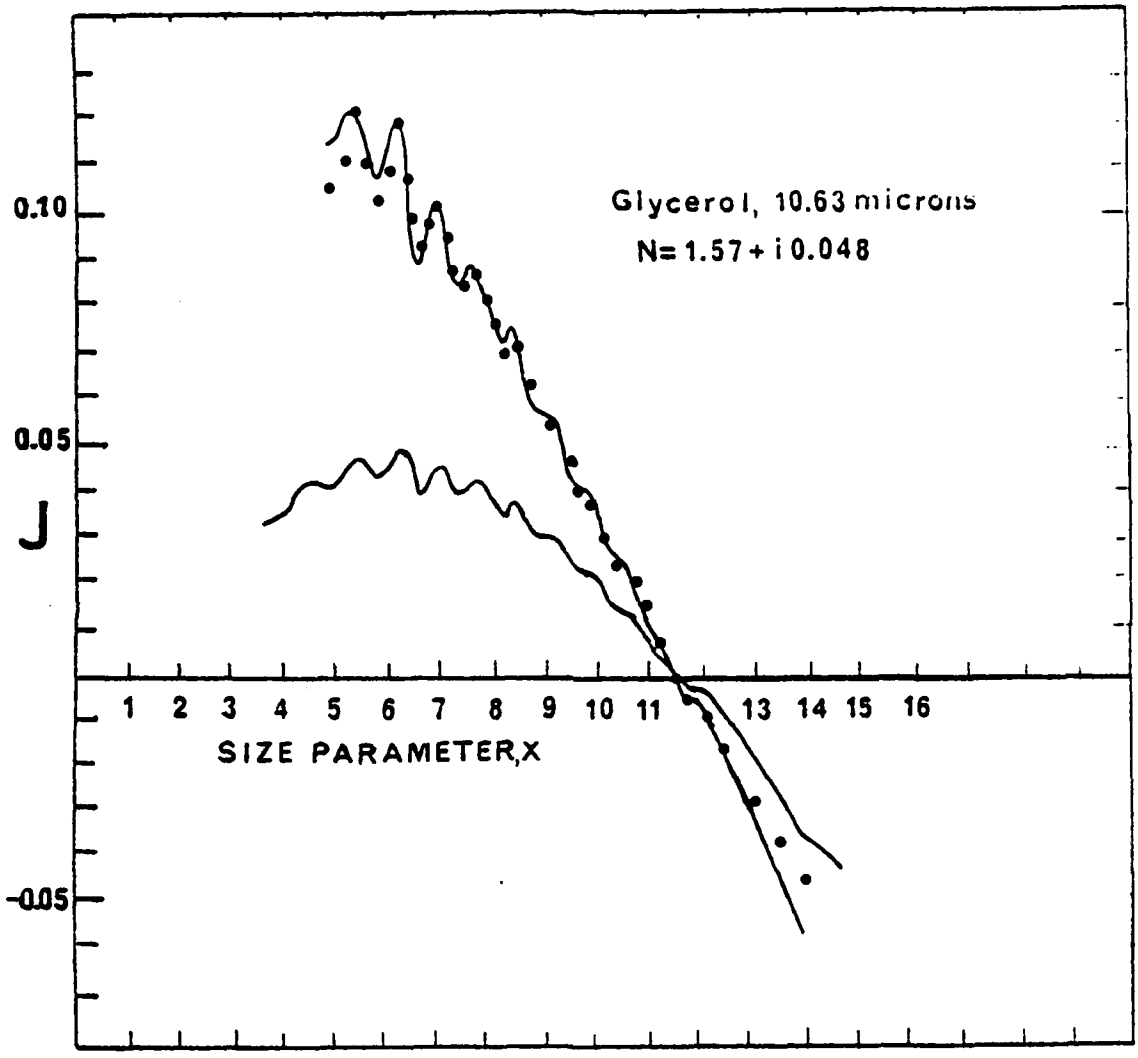


Fig. 6 Evaluation of J of Figure 3 when corrected for the breakdown of continuum hydrodynamic theory. Points show the first principles calculation of J from Mie theory.

In this limit the electromagnetic analysis is simple, and leads to the factor J of Eqn. (2) of the form

$$J = \frac{3}{8}\left[\frac{12\ \rho\varepsilon_r''(\omega)}{|\varepsilon_r(\omega)+2|^2}\right]\left[\frac{2}{15}\ \rho\ \varepsilon_r''(\omega)\ \left(\frac{5/4}{|\varepsilon_r(\omega)+\frac{3}{2}|^2} - 1\right)\right] \qquad (13)$$

where $\varepsilon_r''(\omega)$ is the imaginary part of the relative dielectric constant $\varepsilon_r(\omega)$. This expression has been factored so that the first term in brackets is the absorption efficiency Q_a and the second term is the anisotropy factor A_z. It should be noted that two resonances appear in Eqn. (13). The first is a resonance at $\varepsilon_r = -2$, in the absorption, and the second is a resonance in the anisotropy factor at $\varepsilon_r = -3/2$. The first of these is the resonance associated with the surface dipole mode of the sphere, and the second belongs to the surface electric quadrupole mode. A curious result of Eqn. (13) is that when the surface quadrupole mode is excited, J may change sign and become positive, so that the radiometric force is reversed. This effect, which is caused by the interference between the quadrupole and dipole modes, should occur in Ag at 347 nm, and would render the field at the back of the particle to be considerably larger than that at the front! A typical distribution of these two fields, just below the dipole resonance, is shown schematically in Fig. 7, with the relative phase of the two modes in front and back of the particle as shown. A comparison of the results obtained from Eqn. (13) with a full Mie calculation (i.e., including all modes) shows that Eqn. (13) is valid within 15% for silver particles of radii less than and up to about 100 Å. Thus, the major portion of the effect is due to these two lowest modes alone.

The field reversal predicted in these very small particles was so surprising that we extended the Mie calculations to an 850 Å radius, where experiments are tractable. Fig. (8) shows the results of these calculations at a nearby wavelength of 355 nm. This figure is a topological plot of the square modulus of the internal field within an equatorial slice of the particle. We observe directly the reversal in the internal sources of heat generation proportional to the field intensity. The inset in this figure gives the ratio of the radiometric force F_r to the radiation pressure force F_{rp} (which is always forward), assuming the bulk thermal conductivity of silver. This ratio is just below unity, and suggests that if the thermal conductivity in small silver particles is somewhat reduced, the net force on silver particles can also reverse.

The effect of competing resonances, and radiometric force reversal, is not expected to exist only for surface plasmons in metals. A similar effect should occur in ionic solids when associated with surface phonons (e.g., in SiC, for which $\varepsilon_r = -3/2$ at 10.63 microns). In fact, while the large thermal conductivity of a metal such as Ag limits the absolute radiometric force so that photon pressure is dominant, the effective resonance in SiC at

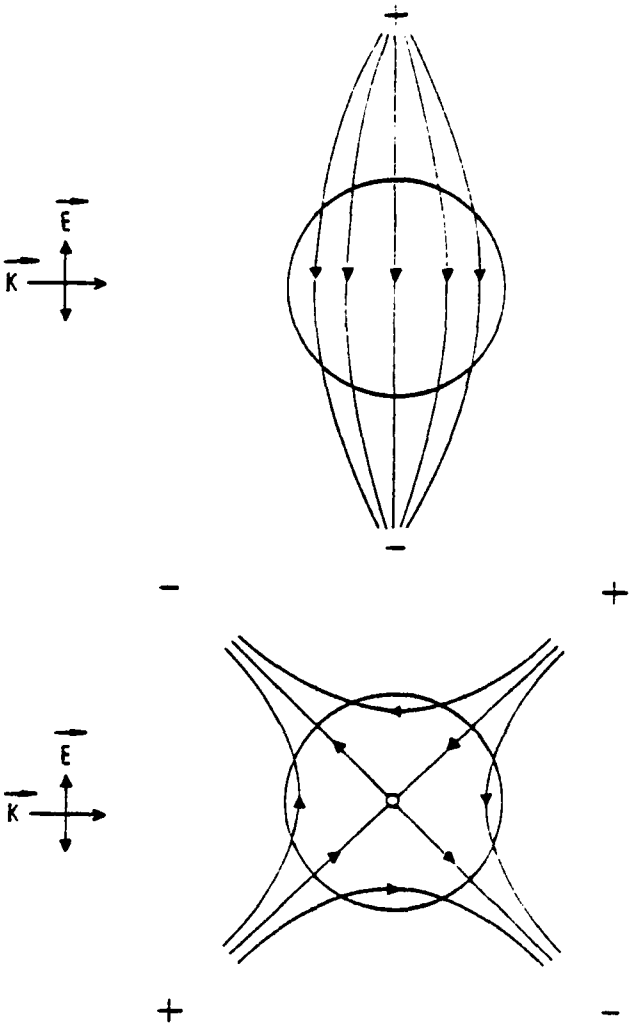


Fig. 7 Dipole and quadrupole field distributions in a small sphere close to the quadrupole resonance. The fields have the same direction at the back of the particle, and opposite directions at the front, producing anisotropy.

infrared wavelengths is sufficiently large so that the overall force is reversed, even assuming unchanged bulk thermal conductivity. It would be interesting to look at semiconductors such as InSb in which the thermal conductivity is less than 1/20 that of SiC, and surface plasmons can be steered in frequency by appropriate doping.

A paper on the influence on photophoresis of the interference of surface modes has been submitted for publication.[52]

We saw above that the anisotropy effect summarized by (13) is due to the presence of a quadrupole mode in addition to the expected dipole mode, and that the anisotropy is most pronounced at or near the quadrupole resonance for the small particle.

In B,(4), we have shown that a simple formulation may be employed to determine the conditions for resonance for spherical particles. This formulation involves spherical transmission line theory, which is a rigorous technique in electromagnetics; a summary that is relevant to our needs here is presented in the above-mentioned section. The simplicity in the formulation is achieved because the method yields the partial information we seek, i.e., the resonances of this class of spherical modes, in a direct manner, without unnecessary appendages.

The general rigorous equation for the resonances of spherical E modes is given earlier as equation (8). It is used in B,(4), to derive directly the well-known result for the lowest, or dipole, mode in the limit of small particle size, showing that the result is, as given by (11),

$$\varepsilon_r = -2$$

To our knowledge, no paper in the literature has so far been concerned, in the small size limit, with other than this lowest mode, which corresponds to mode number n=1. The condition on the value of ε_r for the resonances of higher modes in the small size limit can also be readily obtained from the general relation (8).

In the small size limit ($kR \ll 1$), the expressions for the spherical Hankel and Bessel functions become

$$\hat{H}_n^{(2)}(kR) = \frac{j(n - \frac{1}{2})!\ 2^n}{\sqrt{\pi}\ (kR)^n} \tag{14}$$

$$\hat{J}_n(kR\sqrt{\varepsilon_r}) = \frac{(kR\sqrt{\varepsilon_r})^{n+1}\ \sqrt{\pi}}{(n+\frac{1}{2})!\ 2^{n+1}} \tag{15}$$

When the derivatives of (14) and (15) are taken with respect to their arguments, and substituted, together with (14) and (15) into

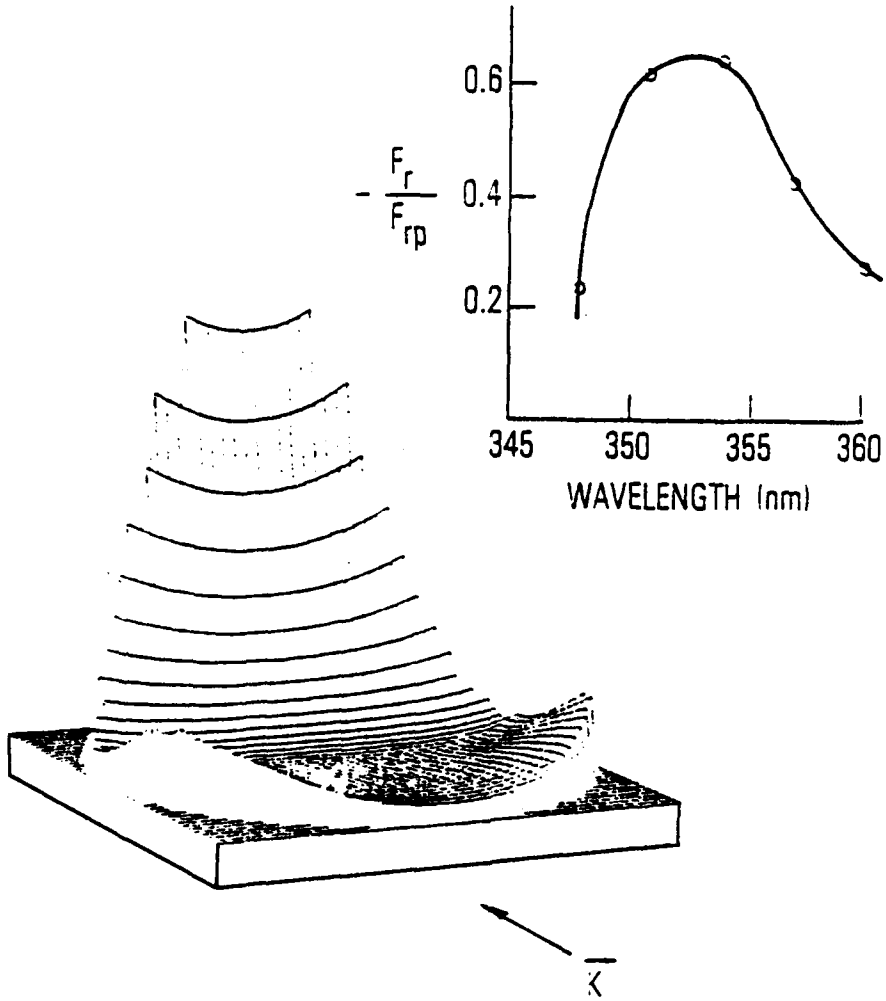


Fig. 8 Front to back distribution of the internal field intensity in a silver particle of 850Å. The inset gives the ratio of the radiometric force F_r to the photon pressure force F_{rp}.

relation (8), one obtains

$$-\frac{kR}{n} = \sqrt{\varepsilon_r}\ \frac{k\sqrt{\varepsilon_r}\ R}{n+1}$$

or

$$\varepsilon_r = -\frac{n+1}{n} \qquad (16)$$

Relation (16) is a generalization for arbitrary mode n of the familiar relation (11) for the lowest mode, n=1. For n=2, corresponding to the quadrupole mode (which is of importance in the explanation of the phenomenon discussed above, involving small particles which can move backwards), we see that $\varepsilon_r = -3/2$ at the resonance. We may also note that for high values of n the resonances all approach the limit $\varepsilon_r = -1$. In addition, since the value of ε_r usually varies rapidly with frequency, the resonances for the n=1 and n=2 modes should occur very close to each other in frequency. In fact, even for silver particles, for which the intrinsic loss is small, it is found that the resonance curves for the two lowest modes overlap each other significantly.

There is an extremely interesting implication in the above results from a conceptual standpoint. It has always been assumed that, in the limit of very small particle size, the electromagnetic scattering is of the Rayleigh type, i.e., the scattered field distribution is that corresponding to a dipole mode. Even in the vicinity of a plasmon resonance for these small particles, the field is still assumed to be of dipole shape, although its amplitude is greatly enhanced by the resonance. We see above, however, that at a quadrupole resonance the scattered field will not be of the Rayleigh type, but instead of quadrupole form or a combination of dipole and quadrupole resonances. Thus, the scattered field due to small particles may be of Rayleigh form over most of the frequency range, and under most conditions, but the exceptions can yield new and interesting physics.

c) Structure Resonance Modulation Spectroscopy

To complement our work on photophoresis we have developed a new spectroscopy for the independent measurement of the absorption efficiency Q_a, which enters as a factor in the expression for J in Eqn. (2). This method, referred to as Structure Resonance Modulation Spectroscopy (SRMS) allows one for the first time to make direct measurements of absorption efficiency on a single isolated microparticle, in the infrared. SMRS utilizes the properties of the natural electromagnetic resonances of a sphere and provides a means for constructing broadband absorption spectra at low intensities.

The resonances of a nonabsorbing sphere of size about 1 micron are seen as extremely narrow spikes in the scattered light

excitation spectrum when the particle is subject to irradiation in the visible.[53] These resonances are so narrow, in fact, that a fractional change of size as small as 10^{-4} can completely shift a narrow resonance past a laser of constant frequency.[49] For a particle of 5 micron diameter this represents a change in radius of 2.5 Å. Furthermore, a small fraction of this change can be detected if the radius is systematically varied and the variation of scattered light is detected with phase-sensitivity. In practice, we have detected changes in average radius below 0.01 Å. If the size change is caused by thermal expansion due to the absorption of, say IR, then an IR absorption spectrum may be constructed by observing the modulation of scattered visible light.

Figure 9 summarizes schematically the elements of the required spectrometer. The particle is held stationary in a quadrupole trap (not shown), and the IR is provided by a GLOBAR plus optical wedge monochromator combination. The visible probe dye laser is positioned in wavelength near a structure resonance, and its elastically scattered light is viewed at 90°.

Our first experiment has been carried out on a 5.4 micron diameter drop of a solution of $(NH_4)_2SO_4$ held in size equilibrium with water vapor. The SRMS IR spectrum of this particle is presented in Fig. 10. The well-known line seen in the figure is the absorption of the $SO_4^=$ ion. The solid curve is a Mie calculation of Q_a for a 5.4 micron particle of a composition consistent with the vapor pressure equilibrium of the solution. The agreement between the curve and the experimental points is very good, and verifies that the SRMS spectrum is the spectrum of the absorption efficiency of Q_a. It is important to note that this particular sample represents the smallest solution mass (90 picograms) for which an IR spectrum has ever been taken.

A paper on Structure Resonance Modulation Spectroscopy has been submitted for publication.[54]

C. Optical Interactions at Surfaces with Controlled Roughness

One of the nearly universal conditions reported for the enhancement of Raman scattering at silver surfaces is that the surface must exhibit roughness. The common interpretation of this requirement is that roughness provides, at least statistically, the local curvature of the surface that produces the conditions for the buildup of optical fields. This buildup, of course, manifests itself in other optical interactions, as well, especially nonlinear ones that depend on higher powers of the incident and reradiated intensities such as harmonic generation, two or more photon transitions, for example. However, there is considerable uncertainty as to the actual extent of the roughness required. Most film samples are used to form a structure of metallic clusters over several hundred Å in dimension,[20] while a minimum of 100 to 150Å rough skin on a smooth metal surface has been found necessary in order to produce enhancement in at least one set of experiments.[21] Generally, the nature of the roughness, especially the character of the minimum roughness that is required to see field enhancements, remains unclear. A better determination of the conditions

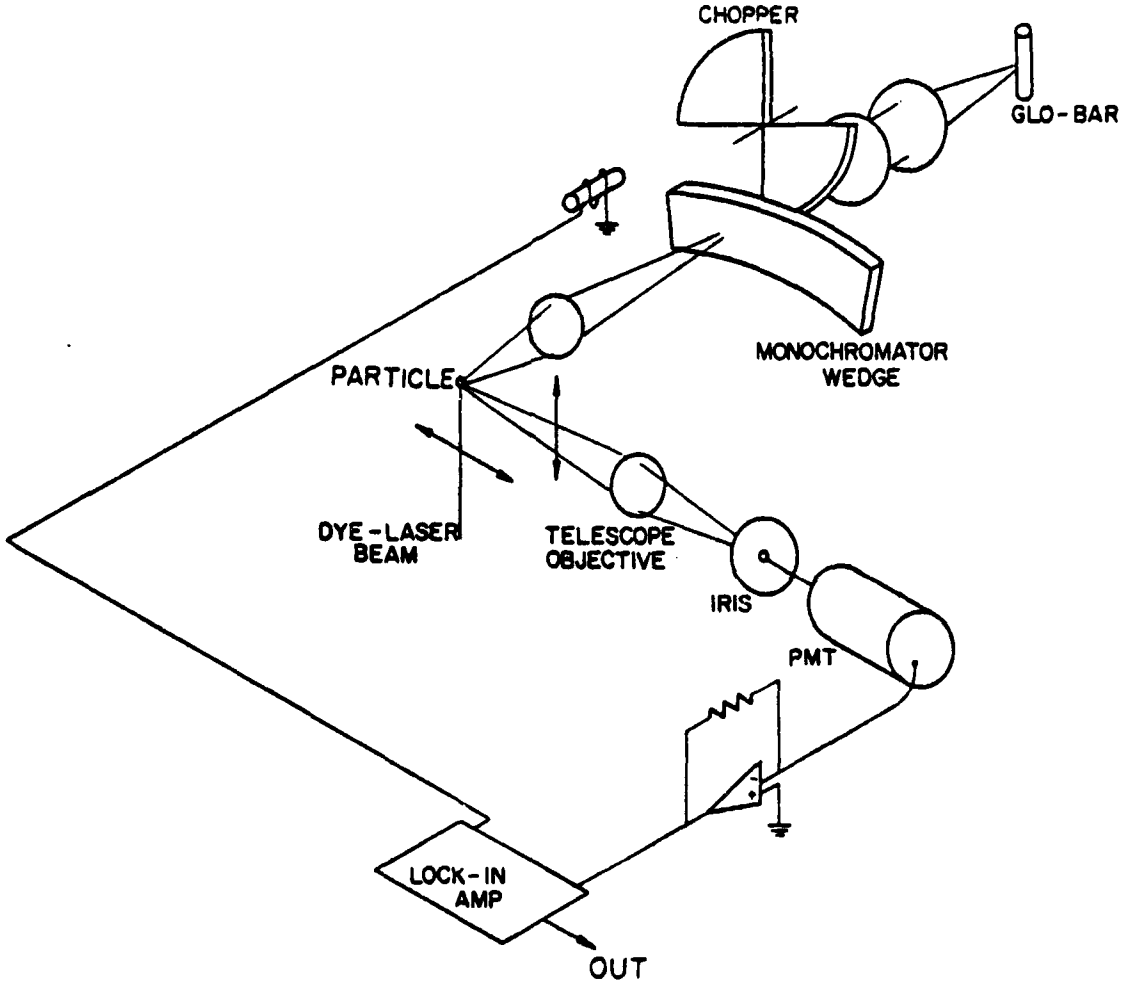


Fig. 9 Schematic arrangement of a Structure Resonance Modulation Spectrometer.

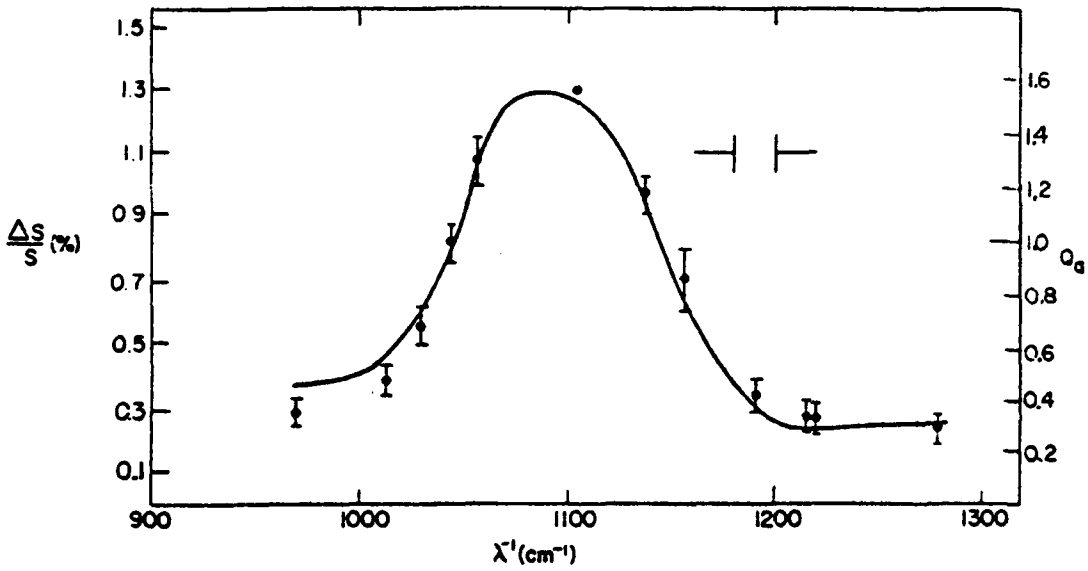


Fig. 10 SMRS result for a $(NH_4)_2SO_4$--water drop. Experimental points: relative change in visible scattered light intensity $\Delta S/S$; full curve: theory of Q_a for this 5.4 micron particle. The horizontal spacer in the upper right gives the bandwidth of the incident IR.

for the onset of field enhancement will contribute to elucidating the extent to which these enhancements result from classical electromagnetic local field buildups alone, and to which other, probably more atomistic-scale, mechanisms also play a role.

We plan to examine this question using an approach that emphasizes correlating the optical responses of a metal surface, most importantly its diffuse scattering and its second harmonic generation, with surface roughness, as this roughness is changed systematically. In another work unit of this proposal (SS2-2), we have outlined a program for characterizing the roughness imposed on an initially very flat and smooth surface by an analysis of how internal electrons scatter from such a surface. In particular, by varying the sample temperature, this roughness can be altered practically at will. Here we propose to correlate the information obtained by such internal electron scattering with the external optical scattering behavior. The first method of measurement is most sensitive for very thin rough layers, but overlaps the range in which the onset of optical field buildup is expected, and it is, of course, ideal for probing smoother surfaces where atomic scale roughness must predominate. In effect, we plan to bring to this problem the full experience and expertise in the production of smooth and controllably rough silver surfaces that is an outgrowth of our earlier and current work under the JSEP program. By emphasizing the nature and control of surface conditions, we expect to clarify some of the contradictions and divergences between different experimental results, and the conclusions based on them, that have appeared in the current literature.

At the same time, we plan to apply the concept of local field buildup to produce optical enhancements to another weak nonlinear interaction which is normally barely accessible. This is the two-photon creation of excitons across the bandgap at insulator surfaces, which can be monitored by looking for the direct radiative decay of these excitons, close to the sum frequency. Such an interaction is an example of possible field enhancement within a solid just below a surface of considerable curvature, and will give information about the details of the dipolar interactions causing optical transitions with such surfaces.

In the newly beginning project on the measurement of the optical properties of metallic surfaces showing a variety of microscopic surface structure, the primary emphasis has been on the development of a portable vacuum system that allows the good preparation and characterization of diverse surfaces, and that can then be brought up to light sources in different frequency ranges located in different experimental setups, so that the sample surface conditions can be maintained. The system makes use of newly acquired pumps and control circuits, and valves. It is beginning to perform satisfactorily, and should be operational shortly.

While the system maintains a good vacuum under equilibrium conditions, we have found that during evaporation the background pressure due to degassing of the metallic source rises to levels that make fast deposition difficult. The surfaces produced under such conditions do not have the purity and the degree of perfection that we used for a definitive measurement. An auxiliary system to handle the gas load during the deposition time is being added.

D. Electromagnetic Theory of Plasmon Resonances on Small Non-spheroidal Metal Particles, Including the Effects of Array and Other Environments

We discussed earlier, in the section on General Background, the fact that there exists disagreement as to the actual cause of the spectacular enhancement of Raman scattering, but that many, if not most, authors seem to believe that the causes are purely electromagnetic. In this portion of the proposed research program, we assume that the enhancement which involves metallic particles is indeed due solely to macroscopic electromagnetic causes, and we show how certain analyses can furnish significant information and insight regarding the enhancement processes, particularly those features such as departure from spheroidal shape and effects due to mutual interaction, neither of which appear in any publications so far.

There are two principal effects, both purely electromagnetic, that combine to produce the remarkable values of enhancement, namely,

(a) surface plasmon resonances on the metal particles, which are dependent on size and shape, and

(b) especially strong concentrations of electric field, which are effective when the excited molecule is properly located.

Effect (b) follows from the specific shape of the particle and also the particular resonance involved. This effect has been termed the "lightning rod" effect by Gersten and Nitzan,[22] and it indicates that electric fields are greatly increased in the neighborhood of regions of strong curvature. Gersten and Nitzan treat only prolate spheroids, but for maximum enhancement they place their active molecules on the axis just above the pointiest part of the prolate spheroid. For a 5 to 1 axial ratio for a prolate spheroid, they calculate an enhancement as high as 10^{11}, in the region of red light, for an optimally-located active molecule. If the metallic particle could be shaped with a very sharp edge or corner (which would involve a departure from the spheroidal shape), the enhancement could be even greater since the electric field becomes divergent at an ideally-sharp edge or corner if the electric field has a component perpendicular to that edge or corner. However, only a few well-located molecules will be treated that well; molecules in average locations will participate in lower enhancements, and will not benefit from the "lightning rod" effect.

Almost all of the active molecules, on the other hand, will benefit from effect (a), since the plasmon resonances increase the fields everywhere on the metallic particle. This portion of our proposed research will deal primarily with effect (a), which contains most of the interesting physics in addition to the unsolved problems.

As mentioned earlier, the surface enhancement of Raman scattering was discovered on rough surfaces of metals, particularly silver. These rough surfaces were mathematically modeled by spheres, or oblate or prolate spheroids, and some agreement with experiment was found. It was shown that the near fields peaked at wavelengths for which plasmon resonances were achieved. These resonances were shape-dependent,

and corresponded to an equivalent dielectric constant for the metal that ranged from near -1 all the way to -10 or more, in the region of red light, for extreme prolate spheroids.

The theories available are only for spheres or for oblate or prolate spheroids, although in general the theories for those shapes are very good. The most complete analysis for spheres was presented by Kerker, Wang and Chew,[23] and that for prolate spheroids was given by Gersten and Nitzan.[22] However, these theories are for isolated particles, and in air (or a homogeneous liquid), not on a silver surface or on a dielectric substrate which can act as a support. In addition, interaction with neighboring particles has not been taken into account in those theories. We may quote directly from the "Conclusions" section of the Gersten and Nitzan paper:[22]

> "The main shortcoming of the present treatment is that it disregards the interaction between the metal protrusions and metal particles lying on the surface. This interaction gives rise to collective electron resonances which were discussed in connection with SERS by Moscovits[24] and by Burstein, Chen and Lundquist.[25] Collective resonances are shifted relative to the corresponding single particle resonance and their presence will affect the present results."

More recently, it was realized that rough surfaces were not efficient since, for a given wavelength, only some of the particles participated in the resonance. It was therefore proposed that a periodic structure be employed, and that the silver particles all be made identical to each other, thus sharpening the resonance and concentrating the effect. Towards this end, Liao et al.[26] presented such an experiment; the structure was fabricated at MIT and measured at the Bell Laboratories. The structure consisted of a two-dimensional array of dielectric (SiO_2) support posts with identical silver blobs deposited on the top of each post; each silver blob resembled somewhat a prolate spheroid with a 3 to 1 axial ratio. Experimentally, they found a resonance-shaped response with frequency, which changed appropriately as the surrounding medium was changed, but the agreement with theory for a 3 to 1 prolate spheroid was mediocre at best. The authors point out that the agreement was not that good because of three limitations in the available theories:

(1) they do not permit departures from the spheroidal shape,

(2) they do not account for the presence of the dielectric support posts, and

(3) they neglect mutual interaction effects between neighbors.

It is clear, therefore, that there is a need for a broader theory that will take these effects into account. In this portion of this work unit, we propose precisely to address the limitations listed as (1) to (3) above.

The general periodic array structure that we propose to analyze and understand is shown in Figure 11. It is a two-dimensional rectangular array of dielectric support posts, placed on a dielectric substrate which can be of similar or different material, and with silver deposits on each of the support posts. We are selective here in choosing the cross sections of the dielectric support posts and the silver deposits to be rectangular, since they are then compatible with the rectangular nature of the array. The solution for the structure then becomes tractable, and in fact we can hopefully even achieve analytical results which are simple enough in form to yield physical insight as well as reasonably accurate numerical answers. In addition, the rectangular shape is a likely result of most of the actual deposition processes.

We analyze the structure in Fig. 11 by choosing a unit cell surrounding a typical support post plus silver deposit. The vertical walls of the unit cell are so-called "phase-shift walls," the relative phase shift between them being dependent on the angle of the incident plane wave and their nature dependent on the polarization of the incident wave.

An approximate equivalent network based on this unit cell and representative of the two-dimensional array shown in Fig. 11 is presented in Figure 12. This simplified version consists of a transmission line representing each of the regions comprising the unit cell; each transmission line is characterized by a propagation constant and a characteristic impedance, and the analytic forms for these quantities will depend on the cross-sectional nature of the unit cell in that region and the angle and polarization of the incident plane wave. The simple form shown in Fig. 12 requires an assumption and contains an approximation. The assumption is that the array spacing is sufficiently small that only a single mode will propagate in the unit cell. This assumption seems to correspond to the physical situations encountered; if an additional mode is present, however, and it couples to the basic mode, then an additional transmission line must be employed. The approximation is that rigorously there exist susceptance elements which occur at the junctions between the transmission lines; these susceptance elements are neglected because we expect on the basis of extensive earlier experience with real dielectric materials that these susceptance contributions are very small. Thus, we believe that the simplified equivalent network shown in Fig. 12 will nevertheless yield accurate results.

The method outlined above automatically takes mutual interactions into account. Since the dielectric support is part of the equivalent network, it is also taken into account. The rectangular shapes involved permit us to handle the solution in an approximate but accurate way, by taking advantage of techniques recently developed by us in connection with studies of open dielectric waveguides.

The basic phrasing of the problem in terms of unit cells and phase-shift walls employs techniques in phased-array antennas which we ourselves have pioneered.[27,28] The application to this problem involves certain important differences from the direct phased array experience, but the basic approach is similar.

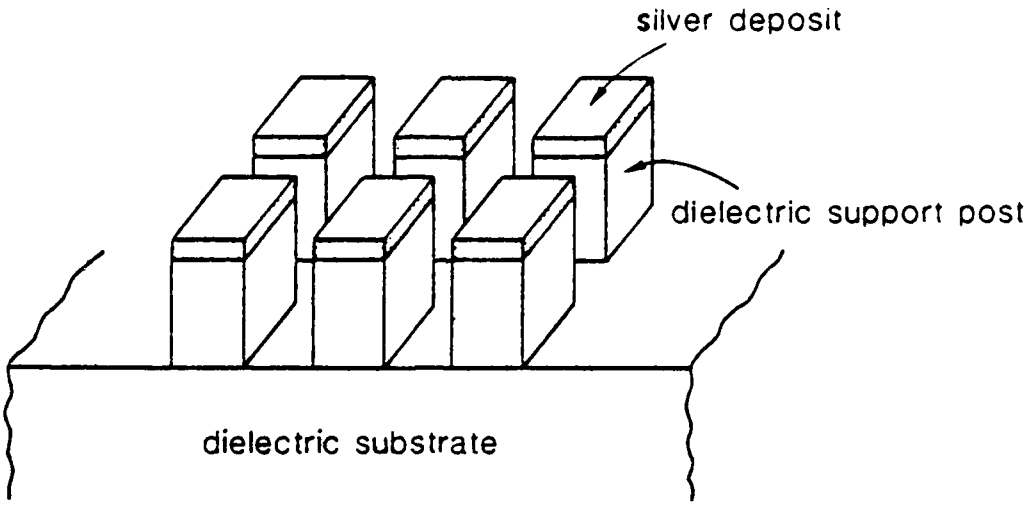


Fig. 11 Two-dimensional rectangular array of silver deposits on dielectric support posts located on a dielectric substrate. This periodic structure is intended to closely model recently-proposed array structures for concentrating the enhancement effects.

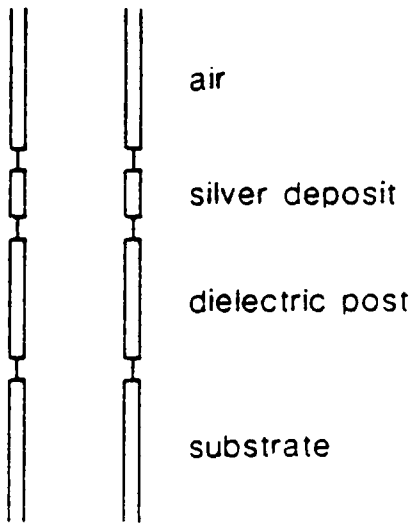


Fig. 12 Approximate but accurate equivalent network of the unit cell that represents the two-dimensional array shown in Figure 11.

Before tackling the periodic array of Fig. 11, which involves several complications simultaneously, we propose to analyze several simpler configurations to develop our understanding more systematically. First, we wish to analyze the isolated parallelopiped shape, corresponding to the silver deposit of Fig. 11, by itself. The analysis would determine the resonances for that structure as a function of arbitrary incidence angle and polarization, and for size and shape dependence. The results would then be compared with available results for "comparable" spheroidal shapes, and approximate agreement would be expected. For example, a cube can be compared with a sphere, a long and thin parallelopiped with square cross section can be compared with a prolate spheroid, and a flat and thin parallelopiped of square shape can be compared with an oblate spheroid.

The parallelopiped geometry permits a greater variety of shapes than the spheroidal geometry does. Thus, these analyses of the isolated structure should yield insight into where the wavelength resonances appear when deviations from the spheroidal shape occur.

The method of analysis to be employed treats the particle as a waveguide with open end terminations, and takes them into account. The incident wave excites an appropriate surface plasmon on the structure, which then resonates at a specific wavelength depending on the geometry and the particular polarization. Use will be made of solutions for surface waves on thin overdense plasma layers originally derived in another context some years ago by Oliner and Tamir.[29,30]

Next, we propose to examine a two-dimensional rectangular array of such parallelopipeds in space to determine the changes due only to mutual interactions. We would employ the unit cell approach and phase-shift walls, but in a simpler context. The mutual interactions would be taken into account automatically by this approach, and we would determine the shift due to the array environment as a function of size, shape, polarization and array spacings. We would see if we could extrapolate the results to predict the shift due to interactions if spheres or spheroids were involved.

As the next step, we return to a single silver particle of parallelopiped shape, but placed on a silver surface. Gersten and Nitzan in their extensive study[22] admit that they did not do this. They took a single prolate spheroid, bisected it, and placed it on a perfectly-conducting plane; they found little change in resonant properties from the full isolated spheroid. The silver surface, as opposed to the perfectly-conducting surface, can itself support a surface plasmon mode. Thus, the interaction between neighboring particles on a silver surface may be greater, and indeed different, from that when the particles are placed on a perfectly-conducting plane. Even the resonances of a single particle will be affected since some energy will be scattered away from the particle in surface plasmon form.

After the single particle on a silver surface, we will analyze a two-dimensional array of such particles on a silver surface.

Then, finally, we treat the complete complex array structure shown in Figure 11. Along the way, we should have developed substan-

tial insight into the new facets under study.

If structures of parallelopiped shape, with sharp edges, of the sort shown in Fig. 11, actually become experimentally important, we may need to take into account more accurately the effects of sharp ends and corners. If so, then a whole new area is introduced, that of edge effects for plasmons on metal surfaces or thin films.

Out of the study as outlined, we will learn the effects of interaction between neighbors, the effect of placing the silver particles on a silver surface, and the influence of the dielectric supports. These effects produce shifts in the wavelengths for which resonances occur, and they affect the magnitude of the enhancement.

In accordance with the proposed research program, we have begun the theoretical electromagnetic studies in Section D by considering the isolated metal particle and the change that would be produced by placing that particle on a metal surface. At this stage, the effort is of course only a few months old. The theoretical analysis for the isolated particle assumes a rectangular parallelopiped shape, as discussed above, and the first stage involves the propagation characteristics of a length of waveguide of rectangular cross section that supports a plasmon solution appropriate to that geometry. The particle is then viewed as a length of such waveguide with open ends. The analytic solution for the waveguide is nearing completion. The remaining important question is how best to treat the open ends in an approximate but accurate way. The final analytical solution should yield the resonances of the isolated particle, and these resonant values will then be compared to those for oblate and prolate spheroids of corresponding size and shape.

A very interesting but simple result has already been found in connection with the metal particle of rectangular shape placed (or grown) on a flat metal surface.

A recent paper[46] by Wood presents experimental data in connection with an added layer of rough silver particles on a smooth silver flat surface. He finds that very thin rough films do not exhibit SERS effects, whereas much thicker films of that type do. He concludes that his experiments disprove the adatom theory of the effect. He observes, in fact, that Raman enhancement occurs only when the roughness features exceed about 150 Å or so. By simple considerations in electromagnetics we can show why the enhancement effects should not appear when the roughness features are too small.

We take as our model a small metal rectangular block placed on a metal plane. A metal half space exists under the metal plane. Similarly, a section of a half space is also present under the metal block, the difference being that that portion of metal half space is slightly higher than the remainder, and that a step junction on each side separates the two regions. Next, we remember that the metal acts as an overdense plasma, so that a surface plasmon is present at the air-metal interfaces when light is shone on the structure. The surface plasmon decays into the air region above and into the metal region below. If the plasmon is tightly bound to the surface, so that the step junction is large relative to the plasmon's transverse (vertical) extent, the plasmon

is trapped in the block region and the Q of the resonance (when it occurs) will be very high. The Raman enhancement will then be significant. If the reverse is true, that is, the step junction is electrically small, the plasmon will simply leak past the step junction, so that the Q of the resonance will be very low and the enhancement negligible.

The decay into the metal region may be characterized by a skin depth δ, given by

$$\delta = \frac{|\varepsilon(\lambda_0)|^{-1/2}}{|\varepsilon(\lambda_0)|} \cdot \frac{\lambda_0}{2\pi} \qquad (17)$$

which represents the vertical distance in which the field has decayed to 1/e of its value at the surface. If the value of δ is large, the presence of the step junction between the block and the flat surface will not be felt unless it is also large.

Calculations for silver as a function of wavelength λ_0 are summarized in Table 1. Interestingly, the skin depth values are almost the same over the whole visible spectrum, being of the order of 250 Å or so for 490 nm light, which Wood[46] used. For step sizes (which correspond approximately to roughness features) much smaller than δ we would not expect significant effects. The threshold for enhancement would be expected to occur for step sizes somewhat smaller than δ, and 150 Å, which Wood found for the threshold, fits well into the predictions of this simple model.

Table 1.

λ_0	$\varepsilon(\lambda_0)$	δ
350 nm	- 2.0	280 Å
490 nm	- 9.1	244 Å
620 nm	-17	232 Å

E. Quantum Mechanical Aspects of Local Electromagnetic Fields Close to Metal Surfaces and Small Structures

Although most of the information available on strong local enhancements has been interpreted on the basis of the Mie theory of the scattering of electromagnetic radiation by spheroidal particles, there are some data that do not fit within this framework.[3] There is a growing concern that other effects are also at work. Some of these, such as the departures from spheroidal shapes and mutual interaction between neighboring scatterers are addressed in Section D above in this proposal. Others are related to the fact that on the scale of molecular dimensions, and similar distances of the molecule above the surface, quantum mechanical effects cannot be ignored. For example, Gersten

and Nitzan, in commenting on the image mechanism for enhancement, state that

> "its true nature has to be investigated with the quantum mechanical nature of the surface, the molecule, and their interaction taken into account."[22]

Similarly, Bergman et al question the use of bulk dielectric constants for very small particles where

> "quantum mechanical effects may become important."[31]

Here we shall be concerned with some of the quantum mechanical effects that can enhance the local fields near a surface to values above those calculated from purely classical electromagnetic theory, along the lines explored by Feibelman for flat metal surfaces.[9] We expect that these effects will be important components of the description of the resonances associated with very small structures and will, for example, be essential in the interpretation of surface photoemission spectra from such structures. This aspect of the proposed research will couple very closely to some experimental work discussed in this proposal. As already implied in the quotes above, and as will become apparent in the discussion to follow, these considerations will introduce another scale of smallness into the discussion, namely the range of quantum mechanical perturbations due to the existence of boundaries, and the distance below which classical modeling breaks down. This scale will in general be superimposed on the scale set by the wavelength of the electromagnetic radiation, but especially for very small structures, which are classically too thin to show interference effects, the quantum mechanical effects may make the dominant contribution to enhancements and other anomalies.

The effects which we shall consider are most clearly seen through a straightforward discussion based on an idealized model of a metal. Consider a free electron gas, confined to a volume by a uniform potential which represents the smeared out positive ionic charge (the jellium approximation). The low energy electronic wave functions inside the volume are approximately described as standing waves, which decay exponentially outside the volume. Due to the Pauli exclusion principle electrons will fill up the lowest energy states up to the fermi surface. The resulting charge density will show a large decrease at the boundaries of the volume, and superimposed on this will be Friedel oscillations[32] with a wave vector that corresponds to that of the fermi surface. Thus there will be rapid variations in the charge density near a surface, which contain a significant quantum mechanical contribution. These rapid variations have an important impact on the properties of applied electromagnetic fields, such that the amplitude and phase of an electromagnetic wave will change dramatically in the surface region. Calculations by Feibelman[9] on a semi-infinite solid have shown that the local fields are enhanced by the rapid variations in charge density, by as much as a factor of ten within the outermost layers of the metal surface, and therefore all photon-surface interactions within this region cannot possibly be treated using classically deduced electromagnetic field amplitudes alone.

In particular, these non-classical field enhancements imply that photoemission arising from the immediate surface layer of a metal is completely dominated by them. The theory predicts a large increase above classical estimates of the photoelectron yield, a different frequency dependence of this yield, as well as substantial changes in the emitted photoelectron's angular distribution because of the rapid changes in the dielectric constant of the solid surface with frequency. Recently, there has been convincing experimental evidence that some of these predictions are correct,[33] so that for the first time there exists a quantitative explanation for the surface photoelectric effect.

Both the predictions of this theory and their experimental verification have been confined strictly to plane metal surfaces. One expects, however, that when the surface geometry is such that classical local field buildups occur, the microscopically correct local fields will show corresponding effects because the self-consistent variations in the charge distribution will also have to satisfy more complicated boundary conditions. It is difficult to predict the order of magnitude of the change in going from a plane to a curved surface, but such a change may in fact be the cause of the recently observed increase in photoelectric yield on small spheres, where classical local field enhancement is insufficient to account for the measurements.[3] At this point it is not clear whether a full-fledged calculation of electromagnetic fields on small spheres is reasonably doable -- the corresponding problem for the field free case remains to be done -- but we will investigate a number of other geometries where the influence of another nearby boundary, or of a simple curvature, becomes important, such as in slightly cylindrical slabs. We will also simulate the effects of surface curvature by studying the interaction of a plane surface with nonplanar incident electromagnetic waves. The success and the results of this work will then dictate whether an effort to describe fields in more complex geometry is warranted. We hope that this work will parallel the extension of the experimentation on single spheres to a size range where the results can be subjected to direct tests.

The general problem of quantum effects on the electromagnetic response of small metallic particles has been approached in terms of formulations for model geometries. The first geometry that has been considered is that of a thin slab. As an initial step this requires the calculation of the eigenfunctions and eigenvalues for the wave functions of the interacting electrons within two parallel boundaries. The eigenfunctions are translationally invariant parallel to the slab's surfaces. The main difficulty is therefore associated with the variations of the wave functions in the direction perpendicular to the two surfaces. For a thin slab, the finite spacing between quantum levels makes the calculation of the wave functions remarkably simple. If the Fermi energy is μ, one only needs to consider a finite number N of bound wave functions, with N given by

$$N \sim \sqrt{2m\mu}\, L/h\pi.$$

The higher energy continuum states that may be needed in addition are reasonably approximated by phase shifted plane waves.

This approach to calculating the needed wave functions bears some

similarity to that used recently by Wood and Ashcroft,[51] although there are some crucial differences. They have included neither the electron-electron interactions nor the finite work function of the slab. Both these features are important in our problem, since they tend to smooth out the variations of the electron density. Since, as Feibelman has shown, it is the gradient of the electron density which controls the quantum enhancement of the electromagnetic fields at surfaces, this smoothing out of the density variations must be known.

The eigenfunctions described above are being used to calculate the electromagnetic response function $\nabla(\underline{r},\underline{r}',\omega)$. The resulting conductivity separates into a local and a non-local part. The local part is given by a Drude-like formula that depends on the local electron density. This density has Friedel oscillations centered about the surfaces of the slab. These are the local variations in the conductivity that Feibelman found to be important in determining the enhancement of the electromagnetic fields at a single plane surface. The non-local response also shows oscillatory behavior. This contribution is not included in the Ashcroft calculation.

We have now obtained the dielectric response, of the thin jellium slab, within the time dependent local density functional approximation. The wave functions were obtained, as by perturbation theory of the starting basis set obtained from the solutions of a finite square well potential. The difference between this potential and the full density functional potential was treated to first order in perturbation theory. The reasonableness of this procedure was assured by comparisons of the density with that obtained by Lang and Kohn for the semi-infinite jellium model of a single surface.

The comparison shows that the perturbation theory for the slab has a slower variation in the electron density near the surface than in the Lang and Kahn scheme. The Friedel oscillation occurs deeper within the slab and the tail of the electron density in the vacuum is larger.

The calculation of the electromagnetic field passing through those slabs has also been carried out. The integral equations for the vector potential were solved for frequencies much lower than the threshold for photoemission. For such processes, only the bound state warefunctions contribute to the electromagnetic response. In this case only a finite number of wavefunctions need be incorporated in the basis state. One of the quantities that is quite simple to evaluate here is the reflectance. This has been studied as a function of the frequency of the incident field. Several resonances are present, below and above the plasmon frequency. The sharpness of these resonances depends crucially on the assumed electron mean free path. The nature of this dependence is more critical than one might expect. A mean free path of even as much as four times the slab thickness can almost completely wash out the sharpness of the resonances. This sensitivity has to be studied further.

The quantum effects on the electromagnetic response of small metal particles are best characterized in the non-local conductivity.

$$\sigma(\vec{r}, \vec{r}; \omega)$$

Systems that are small enough to exhibit quantum size effects have dimensions that are typically no larger than a few hundred Ångstroms. For systems this small, depolarizing fields have an important role in quenching some of the quantum mechanical enhancements. In particular, at very low frequencies, quantum theories have suggested that the infrared absorption should be enhanced over and above the absorption expected on the basis of classical theories. Depolarization fields do, in fact, tend to suppress this expected enhancement.

A particularly interesting set of systems would be those in which quantum size effects may exist in a macroscopic system. A specific example could be random quasi-one-dimensional systems, such as K_2 Pt $(CN_4)_3$ (H_2O) in which the Pt ions form one-dimensional chains. The electronic properties of these systems are dominated by the Pt ions. The dielectric constants of these systems are anomalously large and anisotropic. Typical values of the dielectric constant along the direction of the chains can be as large as 100, and may be varied by intentionally introducing impurities or defects in this system. We have performed a theoretical investigation of the quantum mechanical effects which give rise to this very anomalous infrared response.[55] It appears that the electromagnetic absorption on these systems is composed of many very sharp lines of an entirely quantum origin. The intensity of these is extremely sensitive to vacancies on the Pt sites, very much in analogy with the mean free path sensitivity already mentioned above.

5. NOVEL MEASUREMENT FACILITY FOR SINGLE PARTICLES

The following is a description of our novel facility for determining a variety of properties of small particles, using a photophoretic spectrometer.[11] The versatility of this technique has led to a series of unique methods for the measurement of

1) particle charge Ze
2) particle mass m
3) the complex refractive index N = n + iK
4) particle photoelectron spectra

all on a single particle of micron dimensions, comparable to the wavelength of light. As already discussed in Section B, the same method can also be extended to investigating surface interactions on the single particle.

The facility includes light sources covering the spectrum from the vacuum ultraviolet (0.1μ) to the middle infrared (10.6μ), and it contains the following equipment:

1) D_2 with Li F window and monochromator (0.1-0.3μ)
2) 100W Xe arc with monochromator (0.2-2μ)
3) Ar^+ pumped CW dye laser (0.5-0.6μ)
4) Tunable CO_2 laser (9.4-10.6μ)

All of the unique methods listed above are based on an electrostatic-levitation procedure first discussed by Arnold for the measurement of charge and mass.[34] The diagram for this configuration is essentially

shown in Fig. 2 of Sec. B. It consists of a precise force balance capable of measuring mass equivalent forces of 10^{-14} gm. It is an electrostatic balance in the form of a Millikan chamber, in which the particle position is maintained by utilizing an electro-optic servo system responsive to both changes of position of the particle (A) and the rate of change of position (Bd/dt), that adjusts the vertical field between the capacitor plates through the voltage divider circuit DVM. Optical excitation of the particle from below, by one of the sources listed above, may, for example, change the charge on the particle, or it may induce a photophoretic force because of absorption of the light by the particle. Either effect will cause the apparatus to rebalance in order to maintain the particle's position. We therefore obtain a direct measure of such forces. Although Fig. 2 shows the photophoretic spectrometer in schematic form, the actual system, which is considerably more sophisticated, is detailed in Reference 11. At present, it has a sensitivity of 0.2 electrons, 0.1% mg photophoretic force and a background of less than 1 electron/day.

To indicate the versatility of this configuration we will now discuss some of the measurement procedures for the quantities mentioned above. The number of discrete electronic charges on a particle is determined by causing it to emit a single electron at a time and to then determine the holding voltage before (V_i) and after (V_{i+1}) emission. The simple relationship

$$Z_i = (V_i)^{-1}/(V_{i+1}^{-1} - V_i^{-1}) \tag{18}$$

follows from the Millikan balance equation. The mass of the particle is then determined from

$$m = \frac{(Z_i\ e)\ V_i}{dg} \tag{19}$$

where e is the electron charge, d is the separation between the plates and g is the gravitational acceleration.

Figure 13 shows that such steps corresponding to a single electron have been easily detected.[34] As already mentioned, in the absence of ionizing radiation, the charge $Z_i e$ on a particle can be held constant for a very long time. The facility[1] therefore exhibits extraordinary sensitivity and stability.

The real part of the refractive index of the particle can be determined from the period in the oscillating structure seen in the size parameter dependence of the photophoretic force, as exemplified in Figure 3. The size effect parameter $\rho = (2\pi R/\lambda)$ may be varied either by changing the size of the particle or the wavelength λ of the excitation. Chylek[36] has shown that such structure should have a period $\Delta\rho$ given by

$$\Delta\rho = \frac{\arctan(n^2 - 1)^{\frac{1}{2}}}{(n^2 - 1)^{\frac{1}{2}}} \tag{20}$$

In the example shown in Fig. 3, ρ is changed by evaporation of glycerol material at 10.63μ, and the period of the oscillations at small ρ is $\Delta\rho = 0.72$. Using Eq. (20), this result predicts an index of refraction $n = 1.59$, which is in very good agreement with the bulk value for glycerol, $n = 1.57$. Hence Eq. (20) can be used to determine an unknown value of n, once $\Delta\rho$ is measured.

The imaginary part of the refractive index, K, may be arrived at from the force measurement itself.[37] From Sec. B, Eq. (2), we know that the experimentally determined anisotropy factor J depends only on the parameters n, K and R, so that once n and R are known, K may be determined through a correspondence between experiment and computer calculation. At the value of π for which $J = 0$, the field distribution inside the particle, especially as determined by K for known n, must be such as to lead to uniform heating of the particle.

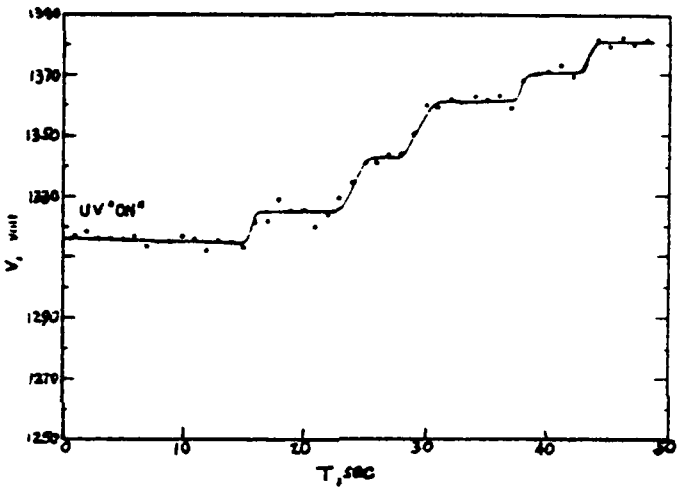


Fig. 13 Holding voltage vs. time for a tetracene microcrystal irradiated by UV. Each small step corresponds to emission of one electron.

As a final example, we show that this technique allows us to learn a great deal about the photoemissive properties of small particles. First of all, the ionization energy may be measured from the onset of electron emission. For a negatively charged particle under balance in our Millikan chamber, the charge is reduced as emission occurs and the voltage necessary to levitate the particle will rise. Figure 14 shows the cube root of the rate of charge loss vs. photon energy for a 10μ tetracene crystallite.[38] It is clear from this figure that the ionization energy is 5.25 eV. In addition to the ionization energy, we can also determine the distribution of photoelectron energies. For a positively charged particle, a coulomb barrier to the escape of electrons will exist. This barrier height is simply q/R, where q is the charge on the particle. If the photoelectron energy distribution is $F(\varepsilon)$, then the yield Y, or the rate of emission, will be proportional to the integral over the electron energy distribution, i.e.,

$$Y(E_b) \propto \int_{E_b}^{(E_p - I_c)} F(\varepsilon)\, d\varepsilon \ , \tag{21}$$

where E_b is the coulomb barrier energy, and E_p and I_c are the photon and ionization energies, respectively. This means that the shape of the electron energy distribution is easily determined from

$$F(q/R) \propto \left.\frac{\partial Y}{\partial E_b}\right|_{E_b = q/R} \ . \tag{22}$$

This electrodeless procedure yields a great deal of information about the photoelectric properties of a small particle. The principal advantages of this measurement are its sensitivity -- we measure the change in charge by as little as a single electron at a time -- and the fact that it is independent of all counterelectrodes that often influence the properties in the surface region of solids.

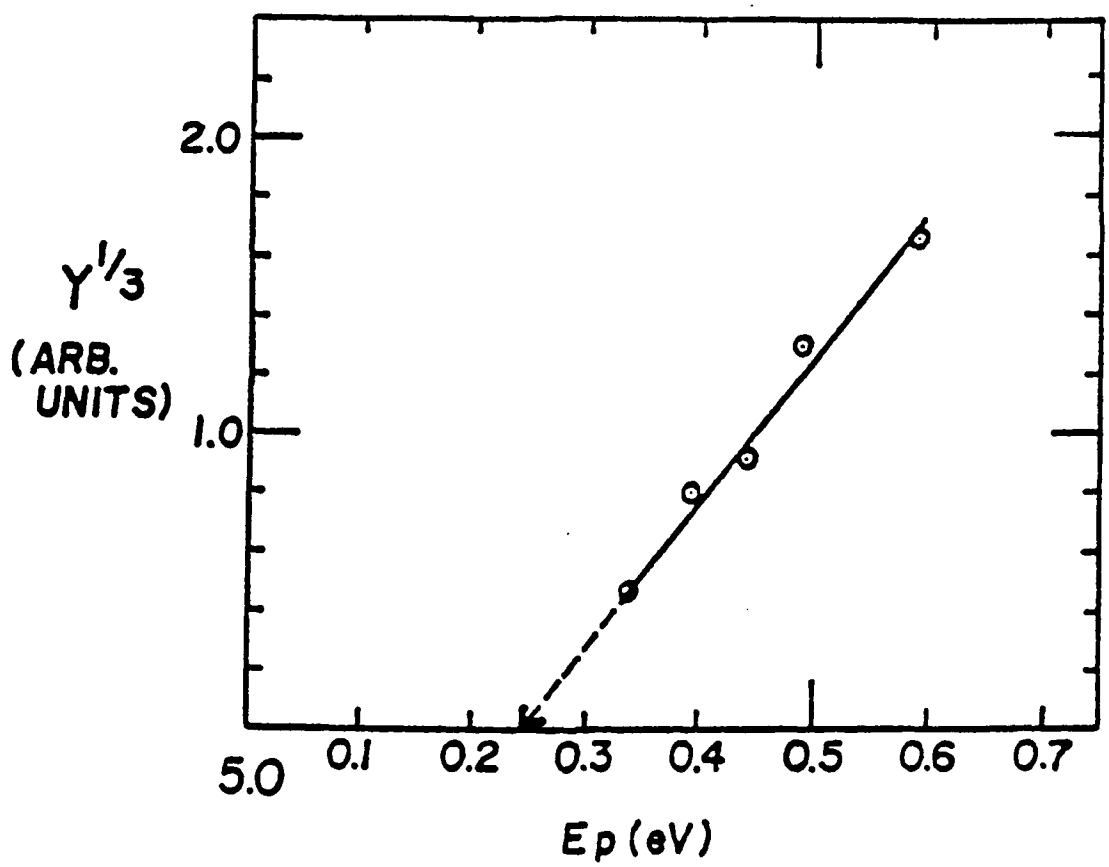


Fig. 14 Cube root of photoelectric yield Y of tetracene as a function of photon energy E_p.

6. REFERENCES AND PUBLICATIONS

1. T.E. Furtak and J. Reyes, Surf. Sci. 92, 351 (1980).

2. Surface Enhanced Raman Scattering, R.K. Chang and T.E. Furtak, editors, Plenum (1981).

3. A. Schmitt-Ott, P. Schurtenberger and H.C. Siegmann, Phys. Rev. Lett. 45, 1284 (1980).

4. C.K. Chen, A.R.B. deCastro and Y.R. Chen, Phys. Rev. Lett. 46, 145 (1980).

5. A.M. Glass, P.F. Liao, J.G. Bergman and D.H. Olson, Optics Lett. 5, 368 (1980).

6. S.L. McCall, P.M. Platzman and P.A. Wolff, Phys. Lett. 77A, 381 (1980).

7. J.O. Gersten, D.A. Weitz, T.J. Gramila, A.Z. Genack, Phys. Rev. B 22, 4562 (1980).

8. R. Dornhaus, R.E. Brenner, R.K. Chang and I. Chabay, Surf. Sci. 101, 367 (1980).

9. P.J. Feibelman, Phys. Rev. B 12, 1319 (1975).

10. A. Otto, et al., "The adatom model, how important is atomic scale roughness?" in Ref. 2.

11. S. Arnold, Y. Amani and A. Orenstein, Rev. Sci. Instrum. 51, 1202 (1980).

12. S. Arnold, to be published in J. Appl. Phys.

13. R.K. Chang, private communication.

14. A.B. Pluchino, Appl. Opt., 20, 531 (1981).

15. F. Ehrenhaft, Phys. Z. 18, 352 (1917).

16. Y.I. Yalamov, V.B. Kutakov and E.R. Schukin, J. Colloid Interface Sci. 57, 564 (1976).

17. G. Mie, Ann. Phys. 25, 377 (1908).

18. M. Kerker, D. Wang and H. Chew, Appl. Optics 19, 4159 (1980).

19. L. Felsen and N. Marcuvitz, Radiation and Scattering of Waves, Prentice Hall, New Jersey (1973), Secs. 2.5 to 2.7.

20. A. Wokaun, J.G. Bergman, J.P. Heritage, A.M. Glass, P.F. Liao and D.H. Olson, to be published in Phys. Rev. B.

21. T.H. Wood and M.V. Klein, Solid State Commun. 35, 263 (1980).

22. J. Gersten and A. Nitzan, J. Chem. Phys. 73, 3023 (1980).

23. M. Kerker, D.S. Wang and H. Chew, Appl. Optics 19, 4159 (1980).

24. M. Moskovitz, J. Chem. Phys. 69, 4159 (1978).

25. E. Burstein, C.Y. Chen and S. Lundquist, Proceedings of the Joint US-USSR Sympos. on the Theory of Light Scattering by Condensed Matter, Plenum Press, New York (1979).

26. P.F. Liao, J.G. Bergman, D.S. Chemla, A. Wokaun, J. Melngailis, A.M. Hawryluk and N.P. Economou, preprint.

27. S. Edelberg and A.A. Oliner, IRE Trans. Antennas and Propag., AP-8, 286 (1960).

28. A.A. Oliner and R.G. Malech, Chapters 2, 3 and 4 of Microwave Scanning Antennas, Vol. II, "Array Theory and Practice," Ed. R.C. Hansen, Academic Press, New York (1966).

29. A.A. Oliner and T. Tamir, J. Appl. Phys. 33, 231 (1962).

30. T. Tamir and A.A. Oliner, Proc. IEEE 51, 317 (1963).

31. J.G. Bergman, et al, Optics Lett. 6, 33 (1981).

32. J. Friedel, Phil. Mag. 43, 153 (1952).
P. Riseborough and D.C. Mattis "Sum Rules in the Kondo Problem," (unpublished).

33. H.J. Levinson, E.W. Plummer and P.J. Feibelman, Phys. Rev. Lett. 43, 952 (1979).

34. S. Arnold, J. Aerosol Sci. 10, 49 (1979).

35. H. Kuhn, J. Chem. Phys. 53, 101 (1970).

36. P. Chylek, J. Opt. Soc. Am. 66, 285 (1976).

37. S. Arnold and M. Lewittes, "Size Dependence of the Photophoretic Force," J. Appl. Phys., 53, 5314 (1982).

38. S. Arnold, M. Pope and T.K.T. Hsieh, Phys. Stat. Sol. (b) 94, 263 (1979).

39. P.S. Riseborough and D.L. Mills, in Valence Fluctuations in Solids, p. 229, L. Falicov, W. Hanke and M.P. Maple, eds., North Holland (1979).

40. P.S. Riseborough and D.L. Mills, Phys. Rev. B 21, 5338 (1980).

41. P.S. Riseborough, Sol. State Comm. 38, 79 (1980).

42. P.S. Riseborough, p. 225, Ref. 39.

43. N.D. Lang and W. Kohn, Phys. Rev. B 1, 4555 (1970).

44. E. Matijević and D.M. Wilhelung, to be published in J. Colloid and Interface Sci.

45. E. Borgarello, J. Kiwi, E. Pelizzetti, M. Visca and M. Grötzel, Nature 289, 158 (1981).

46. T.H. Wood, Phys. Rev. B, 24, No. 4, 2289 (1981).

47. A.B. Pluchino, Appl. Opt. 23, 103 (1983).

48. S. Arnold and A. Pluchino, to be published.

49. S. Arnold and A. Pluchino, "IR Spectrum of a Single Aerosol Particle by Photothermal Modulations of Structure Resonances," Applied Optics, 21, 4194 (1982).

50. M. Lewittes, S. Arnold and G. Oster, "Radiometric Levitation of Micron Sized Spheres," Appl. Phys. Lett., 40, 455 (1982).

51. D.M. Wood and N.W. Ashcroft, Phys. Rev. B, 25, 6255 (1982).

52. S. Arnold, A.B. Pluchino & K.M. Leung, "Influence of Surface-Mode Enchanced Local Fields on Photophoresis," submitted to Phys. Rev. A.

53. A. Ashkin and J.M. Dziedzic, Appl. Opt. 21, 4194 (1982).

54. S. Arnold, M. Neumann, and A.B. Pluchino, "Molecular Spectroscopy of a Single Aerosol Particle," submitted to Opt. Lett.

55. P.S. Riseborough, Z. Phys. B - Condensed Matter 51, 173 (1983).

B. SURFACE STRUCTURAL PROPERTIES OF METALS AND CONDUCTION-ELECTRON SURFACE SCATTERING

Professor H.J. Juretschke

Unit SS3-2

1. OBJECTIVE(S)

To study and characterize the fault structure of the surfaces of nearly perfect thin films of the noble metals, and of very thin irregular overlays deposited on such surfaces at low temperatures in order to understand the mechanisms leading to thermal equilibrium, and the strain distribution in equilibrated surfaces; also to elucidate the very anomalous low temperature behavior of the conductance recently observed by us when the bulk of the thin film is strained by application of a surface charge.

2. APPROACH

To measure internal electron scattering from these surfaces through surface conductivity and its modulation by externally applied surface charge, and then relate the surface scattering to appropriate surface disorder parameters of a new model theory; and to observe internal strains, and their possible low temperature shielding by the strain sensitivity of the electrical conductivity.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

The work during the past year has concentrated on three separate, but related, aspects of the general topic under investigation in this unit.

1) Structural disorder of artificially roughened silver surfaces

The changes of structural disorder or a thin layer of silver deposited at low temperature on a nearly perfect (111) surface of silver have been thoroughly followed through different temperature cycling sequences from liquid nitrogen to room temperature. We have identified the initial low temperature stage of the overlay as characteristic of amorphous metals, and the various intermediate stages at higher T as metastable states of long life, as long as T is not raised further. The transition through these states as a function of temperature appears to follow a critical point description, with a critical temperature T_c of the surface layer around room temperature. This study is now complete, and is being prepared for publication. In addition, a presentation on this topic was made in January 1983 at an American Physical Society meeting.[16]

2) Effective penetration depth of electric fields at metallic surfaces

We have confirmed that the behavior observed at a silver-mica interface, when a thin graphite layer is interposed, also carries over to the silver-vacuum interface, leading to an effective penetration depth of about 15 Å. Attempts to utilize other materials besides graphite to establish a thin layer of poor conducting properties, yet small electrostatic screening length, have had only limited success because of interdiffusion and interface scattering by conduction electrons. Preliminary work using an alumina shows some promise, although the full complexity of a definitive layer arrangement utilizing such a barrier has yet to be tested. A talk on this work was presented at the January 1983 American Physical Society meeting.[17]

3) Strain sensitivity of the resistivity of silver at low temperatures

An experiment to measure this property on bulk silver has been designed and is being tested out. It shows sufficient sensitivity to provide the needed information for interpreting some of the earlier results on the temperature variation of strains induced in thin silver films at low temperatures. Some problems of differential thermal expansion, and of absolute calibration, remain to be resolved.

4. STATE OF THE ART AND PROGRESS DETAILS

Much progress has been made in recent years in exploring the structure of a great variety of materials. The findings confirm that surfaces rarely consist of an abrupt termination of bulk crystal, as if all matter on one side of a mathematical plane cutting the crystal is removed. Instead, atoms in the surface region respond to the absence of nearest or more distant neighbors on the other side of that plane by rearrangements of position and charge distribution. This leads to adjustments ranging all the way from somewhat modified lattice spacings relative to the bulk crystal to major reconstructions that must be considered as new, quasi-two-dimensional solid phases on top of the surface.

Most experimental probes responsible for such description of real surfaces rely on the underlying assumption that, whatever the rearrangement, surfaces remain homogeneous and periodic. This allows converting the usual primary scattering data into information about the surface periodicity as well as the structural arrangements within each periodic surface cell. Deviations from perfect periodicity generally manifest themselves in line broadening and other smearing out of the data, which, if too large, may in fact make the probe ineffectual. In addition, long range deviations from periodicity are often obscured by the restricted resolution and coherence area of the probing beam itself. Such techniques are designed to look for order rather than disorder.

Nevertheless, simple arguments concerning lattice equilibrium lead to the proposition that many of even the most perfect surfaces should show disturbances in their periodic arrangements that are reminiscent of the existence of faults, such as dislocations, in bulk materials. For example, one often finds increases in lattice spacing relative to bulk of the order of one percent between the outermost atomic layers of a surface region. If interpreted as a uniform strain of order 10^{-2}, this

increase should be accompanied by a lateral contraction of the same magnitude (Poisson's ratio), which would then result in a lattice mismatch of one unit cell for every few hundred unit cell displacements along the surface. Actually, though, a strain of order 10^{-2} is far beyond the elastic limit of most materials, and we would therefore expect that, just as in the beyond-the-elastic-limit-response in bulk, the equilibrium structure of the surface will tend to concentrate the strain in small regions of faults, leaving most of the surface strain-free.

The detailed nature of such faults is not clear. They could be vacancies, or line faults, they could cause localized buckling or introduce steps in the surface. For example, surface dislocations were first proposed about 30 years ago,[1] but until recently the level of understanding of most surfaces did not warrant the inclusion of such detail in describing their structure. To the best of our knowledge, there still exist no microscopic or elastic models of the detailed rearrangements for optimum stress relief. But recently, for instance, LEED data on the (111) surface of silver have identified a coherence length associated with the surface of about 75Å,[2] though without further characterizing the details of the structural fault responsible for it. Similarly, the nature of the so-called active sites on a metal surface as possible contributors to various anomalous optical responses of individual molecules attached to a metal surface is receiving increasing attention.[3] Obviously, such faults must also have special importance in the (initial) interaction of clean surfaces with the exterior, ranging from chemical activity, such as corrosion or catalysis, to further crystal growth on the same surface.

For such reasons, basic studies focusing on the structural faults of surfaces rather than on their locally periodic arrangements are timely. Since relatively little is known about this field it makes sense to first concentrate on classes of surfaces that are rather simple, that can easily be brought to thermodynamic equilibrium and are reproducible, that are accessible to probes sensitive to surface faults, and for which an adequate underlying theory exists for their periodic arrangements such that its extension to non-periodic regions can be attempted.

We propose such a study here, formulated with the above criteria in mind. The surfaces will be single crystal surfaces of the noble metals Ag, Au, and Cu, and the probe will be the conduction electron scattering by the surface, as manifested by the surface-sensitive contribution to the electrical conductance of thin films of these metals. The proposal is an outgrowth of a portion of Work Unit SS1-4 in the current program, and builds extensively on the expertise and the information gained there. The noble metals were chosen because, by a fortunate coincidence, they can be made to form reproducibly extremely flat and well-ordered surfaces by properly controlled epitaxy on very thin ruby muscovite mica, and because much recent information is available about the theoretically expected charge redistribution and lattice distortion in the surface regions of these metals.[4]

Surface conductance in thin films becomes an obvious probe, because, first of all and in contrast to most other surface probes, it is

most sensitive to the disorder, rather than the order of the surface. Secondly, in very thin films surface scattering can be made to dominate all other scattering processes contributing to the film resistance, especially at low temperatures. Since modern measuring techniques for conductance have a sensitivity of a part per million, extremely small changes in surface scattering can be detected. For example we will describe below that a single monolayer of disordered material added to an ordered surface produces changes in resistance at liquid nitrogen temperature of more than ten percent. Extrapolating, one should be able to see as little as 10^{-4} or less of a monolayer of added material, and a recent study has, in fact, verified this possibility.[5] In addition, modulation techniques, such as by applied surface charge, can be used to provide additional information and to extract further detail at the 10^{-6} relative signal level. The high sensitivity of this probe, if thought of as a scattering technique, obviously arises from the enormous intensity of the incident beam of charge-compensated and effectively non-interacting electrons. Its chief drawback is that the incident and scattered beam directions are ill-defined, and only their averages can be obtained. However, for purposes of this investigation, which concentrates on surface disorder, the averaged cross-sections obtained in this manner will still contain much information. It is, of course, equally important to possess suitable theoretical models of surface scattering for processing the information provided by experiment. As described below, one such model has recently been developed by us as part of the current research unit, and its initial applications have yielded new and fruitful insights into surface structure.

The unique combination of having available high quality surfaces, in a configuration allowing the use of a very sensitive measuring technique, and backed by a promising and relatively straightforward theoretical model, forms the basis of this proposal.

While the use of resistance to monitor surface reactivity has a long history,[6] recent work along lines similar to those we propose is being reported by groups at CNRS-Orsay, France,[5] and at Kharkov University, USSR.[7] However, both programs are only using resistance measurements, while we expect that the simultaneous information provided by surface charge modulation is important for a more complete surface characterization. The direction of recent surface modulation studies at Clemson University[8] is different from ours, and aims at understanding this effect in semimetals where complications of band structure, anisotropy, and small Fermi level have to be sorted out. An important additional difference between our proposed work and that at Orsay is that we are concerned about absolute information on surface structure as well as any changes in it introduced by small amounts of added-on material. Furthermore, our interpretation makes full use of the fact that at low temperatures and in good films, many successive scatterings off the surface under investigation, as well as the opposite film surface, must be taken into account, rather than using the asymptotic formulas of size effect theory that are valid only in the limit where the film thickness is much larger than the mean free path.

In order to understand the results of our most recent work, it is necessary to place them in perspective by including some immediately preceding stages of progress. The most recent results are included below in their appropriate contexts.

The overall aim of the past and current program was to gain a better understanding of the origin of the scattering of conduction electrons at metallic surfaces. After initial work[9] had demonstrated that the combined information of both the electrical conductance and its variation with surface charging, measured over an extended temperature range, could be satisfactorily interpreted in terms of a surface specularity and its modulation with surface charge, the experimental work split into two separate but interacting directions, both aiming at elucidating the origin of this surface specularity. One direction concentrated on clarifying the various new effects discovered during the study of the charge modulation of the surface scattering of well-defined interfaces, and the other focused on methods of controllably producing surfaces of different specularity.

With respect to the charge modulation of metallic thin film conductance, we were able to show for the first time that the two major responses, one proportional to the first power and the other to the second power of the surface charge density, were primarily associated with surface and bulk properties, respectively, of the thin film sample.[10] Further, we demonstrated, also for the first time and by two independent methods, that the quadratic term was related to elastic strains set up throughout the thickness of the sample by electrostatic charging. In one method, we observed the change in the direction of magnetization (of the order of seconds of arc) with surface charging in permalloy films, caused by magnetoelastic coupling.[11] In the other, we were able to suppress the quadratic signal altogether by efficient mechanical clamping of the sample.[10] This quadratic effect, though not directly concerned with surface properties, nevertheless had an important bearing on understanding the linear, surface-related term. The similarity of the two magnetization-dependent responses in permalloy, one linear and one quadratic in surface charge density, practically forced us to interpret the surface-related signal as also coming from strain-mediated changes in direction of the magnetization originating in the immediate surface region.[12] This discovery set the stage for a serious probing of the proposition that most, if not all, surface scattering of conduction electrons is related to localized strain fields emanating from the surface and extending inward over a small distance. Our most recent progress is based entirely on this interpretation.

However, in order to implement an experimental program aiming in this direction, it had to be reoriented towards studying the effects of charge modulation on free metal surfaces rather than on a metal-insulator interface. In both configurations the nature of the observed effects has been found to be essentially the same, but the free metal surface has the advantages of being simpler, and also of being more accessible, even though the overall experiment itself becomes considerably more complex. In addition, the temperature range of our measurements was extended downward to less than 2°K in order to enhance the effect of the surface on transport properties as much as possible. Finally, to support the interpretation of these measurements, a new theoretical model of conduction-electron surface scattering was developed, based on scattering in depth within a thin surface layer that has electrical properties differing from those of the underlying material of the metallic film.[13] The two parameters characterizing this layer are its thickness, and its mean free path.

The most important findings, interpreted in terms of and fully compatible with our theoretical model, are:[14]

1. The mean penetration depth of surface strains on our well-annealed metal surfaces is about 10 to 15Å, as determined by the thickness of a poorly conducting overlay, such as graphite, that suppresses the linear term in charge-induced scattering (without at all affecting the quadratic term). The decrease of signal with graphite thickness is shown in Figure 1.

2. The surface mean free path of our samples is temperature-independent, and of the order of several hundred angstroms. This is determined by analyzing the (very pronounced) temperature dependence of both the conductance and its modulation with surface charge in terms of our theory. The agreement with charge-modulated scattering and the theory is shown in Figure 2.

3. Detailed application of the theory interprets the surface mean free path as a correlation length for surface order, and thus it also specifies a mean distance between localized fault regions.

4. Both the linear and quadratic responses show anomalous behavior at temperatures below 12°K, with the signals tending to disappear. This is most pronounced for the quadratic term which drops precipitously, sometimes below the noise level, below 7°K, as shown in Figure 3. At this writing, we do not know whether this effect arises from a suppression of strain-sensitive scattering, or from a shielding of charge-induced strain from the bulk of the sample.[15] Either interpretation would have to invoke new and so far never-discussed properties of metals or charged metal surfaces at low temperatures.

According to the first finding, the distance below a metallic surface within which electric fields applied to the surface can affect electrical conduction is apparently much larger (~ 15Å) than the distance of the electrostatic screening (~ 0.5 Å) of the electric field itself. This conclusion was reported last year, and was based on studies of the influence of a thin layer of graphite above the metal on the surface scattering of current carriers in the presence of an electric field applied to the surface of a thin film of silver. The typical decay of this influence as a function of graphite thickness is shown again in Fig. 1. Carbon has the property of covering the silver extremely uniformly, and it undergoes little diffusion into the metal. In parallel with the silver, it does not contribute measurably to electrical conduction, nor does it induce scattering at the silver-carbon interface, as deduced from the fact that there is no change of electrical resistance upon carbon deposition. Its conduction electron density produces a screening depth of about 2 Å to 5 Å, considerably less than the 15 Å observed in Fig. 1, which must therefore reflect non-electrostatic, and presumably elastic, effects.

In order to further understand this phenomenon, we have extended these studies to other layer structures, with screening lengths larger

than that of graphite. First of all, we have confirmed that the results of Fig. 1 apply whether the carbon is between the silver and a dielectric, or when it is on the free surface of silver. This latter possibility is important because this geometry allows a direct examination of the carbon overlay, such as by an Auger probe. We have detected only trace amounts of silver above or through the thin carbon layer, and certainly much less than the amount needed to reproduce the data of Fig. 1 on the basis of an inhomogeneous coverage alone. The penetration depth of ~ 15 Å persists, and the insensitivity of the Maxwell stresses to the presence of the surface layer, reported earlier for a silver-dielectric interface separated by a graphite layer, continues to exist at the free surface.

The extension of this approach to other possible metallic layers of intermediate screening depths has not been too successful. Obvious candidates, such as antimony, give complicated results because they interdiffuse readily, and because they also produce noticeable interface scattering. Attempts to separate the current carrying silver film from a thin metallic overlay by an insulating barrier, such as Al_2O_3, show promise, but may have to be produced and measured entirely at low temperatures in order to minimize diffusion through the barrier in either direction. In addition, it still has to be established how uniform a metallic overlay can be produced on top of the alumina barrier. It may very well be that graphite is unique in satisfying all the requirements of this particular experiment. In that case, it becomes important to devise a corroborative experiment for the screening length of our graphite layer before the interpretation of Fig. 1 can be firm.

In consonance with the work outline of the proposal, and in anticipation of the need to extend further experiments to very low temperatures, we are constructing a new experiment for measuring the strain sensitivity of the resistivity of bulk metals at low temperatures. Information on this behavior does not exist in the literature, and since this behavior is needed to interpret the conductance modulation data due to Maxwell stresses at low temperatures, we have undertaken to design an experiment to determine the needed experimental data.

We have now established that there is sufficient sensitivity to measure this effect reliably with the same apparatus used to detect the surface conductance modulation of silver films. The major remaining design problem before embarking on the measurements themselves concerns the differential thermal contraction of various components of our sample holder and sample. Since one of the objects of carrying out these measurements is to establish the onset of higher order terms in the strain sensitivity of resistance at low temperatures, we must at all times know the operating point around which the sample is being strained. As this changes with temperature if there is a substantial differential thermal contraction, we are devising another set of piezoelectric controllers to give information on, and, if necessary, to independently vary, this operating position.

The calibration of all piezoelectric transducers over an extended temperature range also remains to be established in order to put the

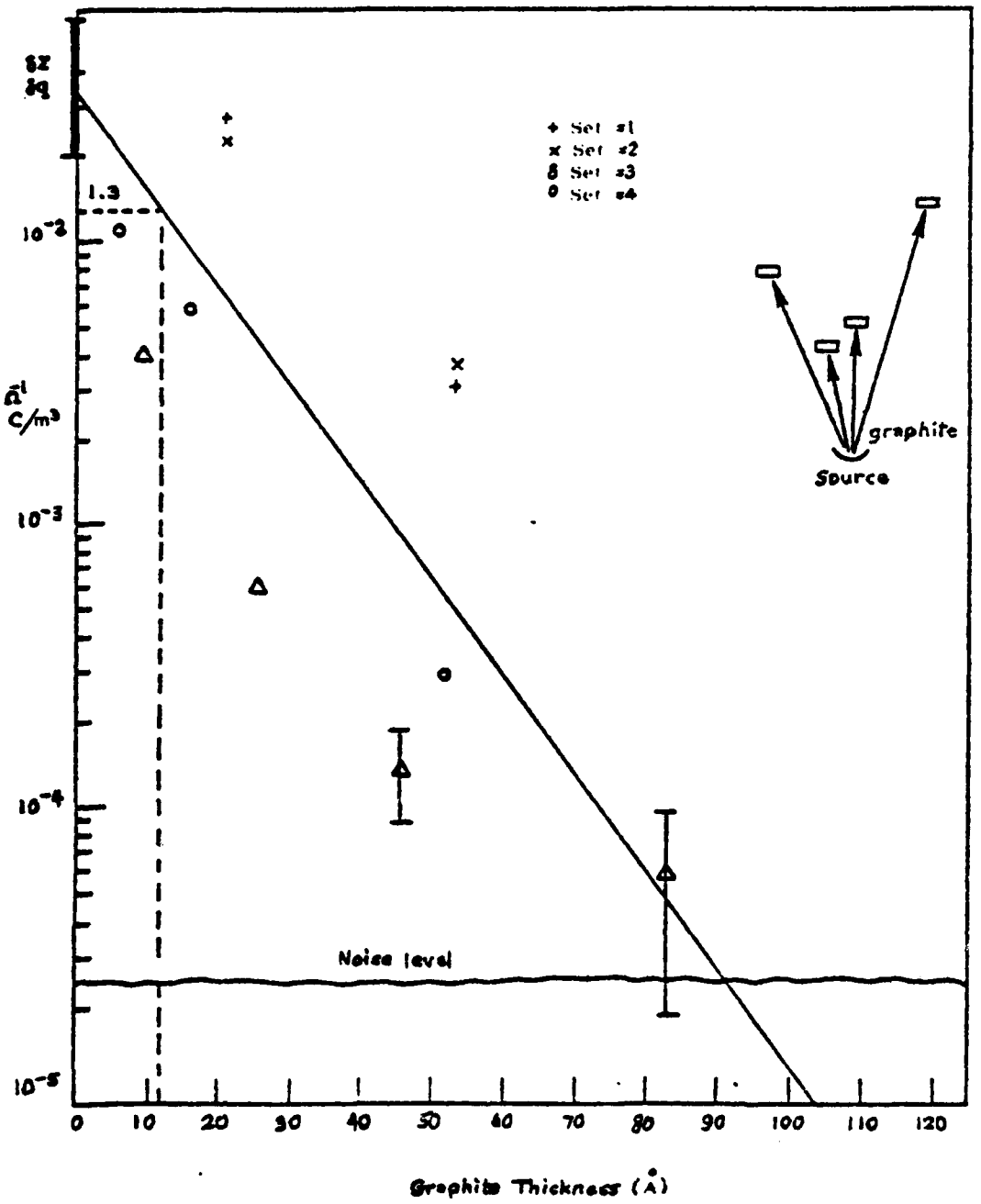


Fig. 1 Variation of surface charge density modulation of silver film conductance, $\delta\Sigma/\delta q$, with thickness of graphite overlay, at a silver-mica interface.

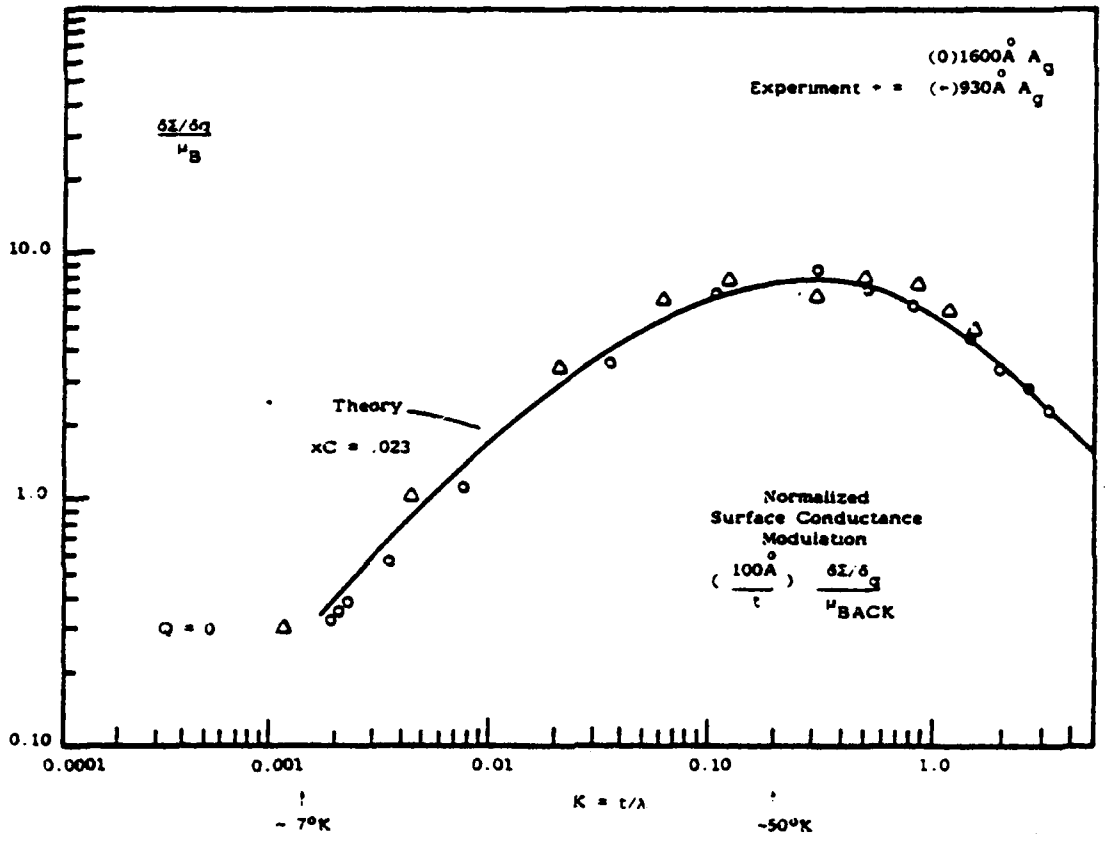


Fig. 2 Temperature dependence of $\delta\Sigma/\delta q$ at the free metal surface of two silver films, relative to the bulk mobility μ_B, and reduced to a 1000Å thickness. The horizontal scale, in terms of the universal size effect parameter K (= thickness/bulk mean free path), is directly proportional to the absolute temperature T. The experimental points follow the line based on the theory of surface scattering caused by surface charge-induced modulation of surface disorder.

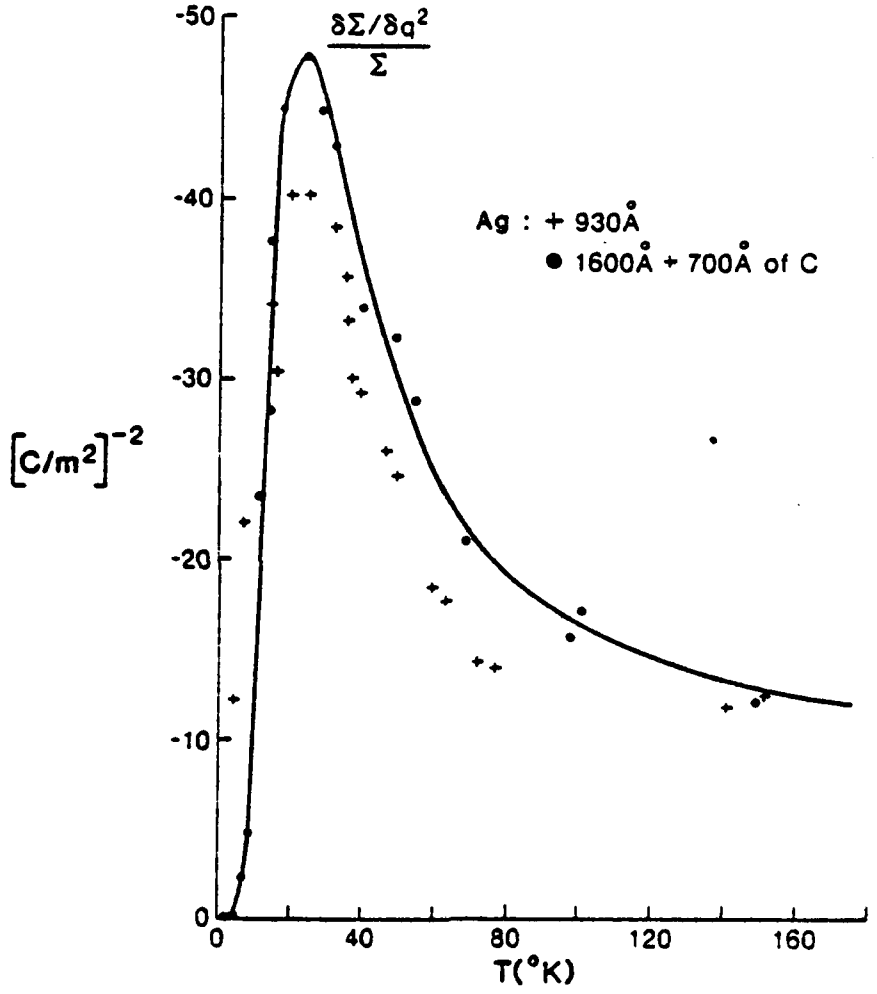


Fig. 3 Relative film conductance modulation proportional to q^2, as a function of temperature for two samples: a 930Å Ag film charged on its free surface (+), and a 1600Å Ag film charged on the outside of a 700Å graphite overlay (·). Note the anomalous peak followed by a steep drop at very low temperatures.

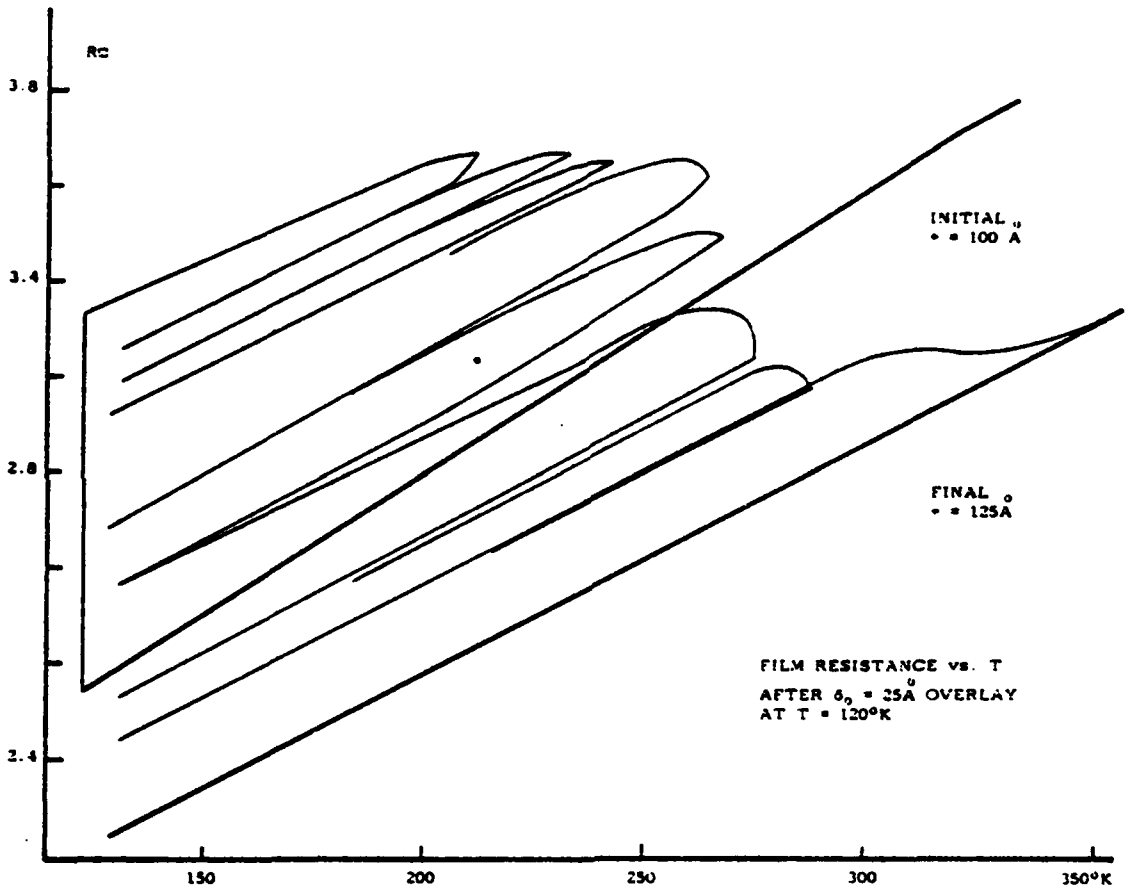


Fig. 4 Resistance vs. temperature of an initially well-annealed 100Å film of silver that is then covered with 25Å Ag deposited at T = 120°K. Each straight line portion at low temperature represents a stage of partial surface layer annealing, such that along this line R is reversible below a temperature T_c characteristic of that line.

observed effects on an absolute basis.

With respect to producing surfaces of controlled specularity, we have refined the method based on low temperature deposition of thin metal overlays on films with initially well-defined surfaces.[9] By regulating the temperature variation to produce partial annealing of the overlay, we can now obtain a practically continuous variation of effective surface specularity.[16] (See Figure 4.) The initial resistance, after depositing an overlay of 25 Å at T = 130°K, increases by more than 30% and the annealing process of the overlay is broken up into seven intermediate stages. Each stage has its resistance purely reversible below a temperature T_c, but shows an irreversible decrease to the next stage when the sample is taken over T_c. A preliminary analysis of the curves in Fig. 4 in terms of our theory of surface scattering yields an effective surface layer thickness and the order correlation length ℓ_i as a function of T_c, as shown in Fig. 5. Both are strong functions of annealing, and both stabilize at finite values for a fully annealed thermodynamically stable surface. By rapidly increasing the sample temperature above a given T_c, and then holding it there, we have recently started to monitor the slow component of the change of surface scattering with time, (Fig. 6), as the overlay scattering approaches the value appropriate to the new T_c. The longer delay time at higher T is indicative of a higher activation energy for diffusion. A detailed analysis of these results remains to be carried out, and only after the initial part of the curves has been corrected, where necessary, for thermal lag between sample and nominal ambient temperature.

We have made substantial progress in determining the structural disorder of thin metal overlays deposited on nearly perfect surfaces of the same material at low temperatures, by measuring the surface scattering of conduction electrons. Any interpretation of such data has to take into account that even a well-annealed surface still shows a residual surface structural inhomogeneity characterized by a depth of 10 to 12 Å, and a lateral correlation distance of about 400 Å, both independent of temperature. We have made use of this fact to find the correlation length within the rough overlay as a function of temperature. By using overlays also of thickness 10-12 Å we assure that throughout the annealing process of the overlay the effective layer thickness does not vary significantly, so that all observed changes in surface scattering are primarily attributable to changes in the lateral correlation length ℓ_i. Figure 7 exemplifies these changes in ℓ_i as a function of temperature for 12 Å overlays of silver on epitaxial silver films varying in thickness between 300 and 750 Å. It shows clearly that this length variation is independent of film thickness, and that atomic diffusion in the overlay commences at about 220 °K, and is completed around 310 °K. Both of these temperatures are significantly lower than in bulk. Our progress is clearly apparent when comparing Fig. 7 with Fig. 5, based on preliminary measurements. Once the data of Fig. 7 are established, they can then be used to measure the effective thickness of overlays of different initial thickness as a function of temperature. Figure 8 gives an example of this variation for overlays initially about 6 and 20 Å thick. As expected, they give the initial overlay thickness at very low

temperatures, and end up at high temperatures at the common disturbed layer thickness for annealed surfaces. The temperature variation in between the extremes measures the composite effect of the annealing of the overlay, as well as of the surface layer of the original surface.

Our program for investigating the structural properties of these artificially roughened surfaces has relied on determining ℓ_i as a function of various parameters, such as the overlay thickness, the thickness of the underlying structure, and, most importantly, the temperature. This determination, in turn, depends on extracting ℓ_i from measurements of the influence of the surface on electrical conduction. For this purpose we have used a transport theory based on a phenomenological model. Because this model contains, in addition to ℓ_i, a number of other parameters, the measurements program must be such as to overdetermine these parameters, in order to assure that their values are fairly unique, reproducible, and compatible with all aspects of this model. Such a program, testing various independent predictions of the model, has been carried out. It has confirmed that the model is indeed consistent with a great variety of different measurements on the same system, and we are confident that the parameters so obtained represent real physical properties that are probably model-independent.

As an example, Fig. 9 shows the change in electrical resistance at 140°K of a nominally 300 Å thick film as three overlays are deposited, one terminating at Δ = 6 Å, the next at 12 Å, and the third at 20 Å. The continuous curves in Fig. 9 show the fit to the model with a common order parameter ℓ_i = 20 Å. It is seen that the fit not only reproduces the overall magnitude of the change of resistance with roughening, but also the position of the maximum, as well as the slopes of the initial rise and the decrease beyond the maximum. The changes in these results observed when the experiment is repeated on films of larger initial thickness are also consistent with the same ℓ_i = 20 Å. Hence we obtain both a measure of the adequacy of our model to explain such features of the experiment, and a good number for describing the order in the overlay.

The extension of this approach to obtain the temperature variation of ℓ_i as the film plus its overlay are heated at a constant rate has already been demonstrated and discussed (see Figs. 7 and 8), where it was also shown that ℓ_i increases with annealing temperature, but that it retains this value below this temperature if the thermal cycle is reversed, so that annealing is irreversible.

The significance of ℓ_i = 20 Å of the as-deposited overlay is that it agrees with the mean free path measured in amorphous metals; and the irreversible increase of ℓ_i with temperature suggests a continuous series of transformations in the overlay from the amorphous to the crystalline state of the underlying material. The new feature of these results is that such behavior has never before been found in pure bulk metals. These transform spontaneously into crystals immediately upon deposition, at any temperature, and an amorphous state requires stabilization by

impurities. Apparently a disordered overlay of pure material on a crystalline surface of the same material represents a new possibility for maintaining a (very thin) amorphous pure metallic phase at sufficiently low temperatures.

Since apparently it has not been seen before, there is no theoretical model for such a metastable phase, or its transition to the crystalline state. Rather than construct such a model at this time, we have concentrated on other experiments that might give additional information on the characteristics of this system. These experiments, primarily based on time and heating rate studies, have shown that there seem to exist at least three time scales in any one step towards crystallinity initiated by raising the temperature a few degrees. The first one is of less than one minute, during which most of the eventual total change to the new state is completed. The next one, of about an hour, sees the completion of the remaining change, and, finally, for times larger than this practically no further change takes place. The same sequence is repeated if the temperature is raised once more to a new value. This behavior, all taking place at temperatures quite low for ordinary diffusion, points to a mechanism combining short-circuit diffusion (presumably because of the presence of the free surface) that is stopped by high barriers that are very sensitive to raises in temperature. At the moment, there exists no detailed information regarding the origin of either of these rate processes.

Since the solid-solid transition towards crystallinity must be of order higher than one, one can also attempt to interpret the data of $\ell_i(T)$ of Fig. 7 in terms of a critical point formulation. This would take the form $\ell_i(T) = L(T_c - T)^{-n}$, at least in the range of ℓ_i in which the surface bulk equilibrium does not inhibit the completion of the annealing of the overlay. Figure 10 shows that such a functional relationship is indeed obtainable for thin overlays (curve 1). Furthermore, a similar behavior is found for the annealing of the bulk of films deposited in a manner similar to the overlay, but, of course, the critical temperature T_c in this case is higher (curves 2 and 3). In all curves, the exponent n is of order 2. Both this value, and the magnitude of T_c in bulk and for the overlay, must await a detailed model for an explanation.

The above results constitute part of a thesis[18] submitted by Y. Amani for the Ph.D. degree, and are now being prepared for publication.

5. REFERENCES AND PUBLICATIONS

1. C. Herring, "Surface tension as motivation for sintering," in The Physics of Powder Metallurgy, McGraw-Hill, New York (1951).

2. D.G. Welkie, M.G. Lagally and R.L. Palmer, "LEED study of the surface defect structure of Ag(111) epitaxially grown on mica," J. Vac. Sci. Technol. 17, 453 (1980).

3. A. Otto, I. Pockrand, J. Billman and C. Pettenkofer, "The ADATOM Model: How important is atomic scale roughness?"

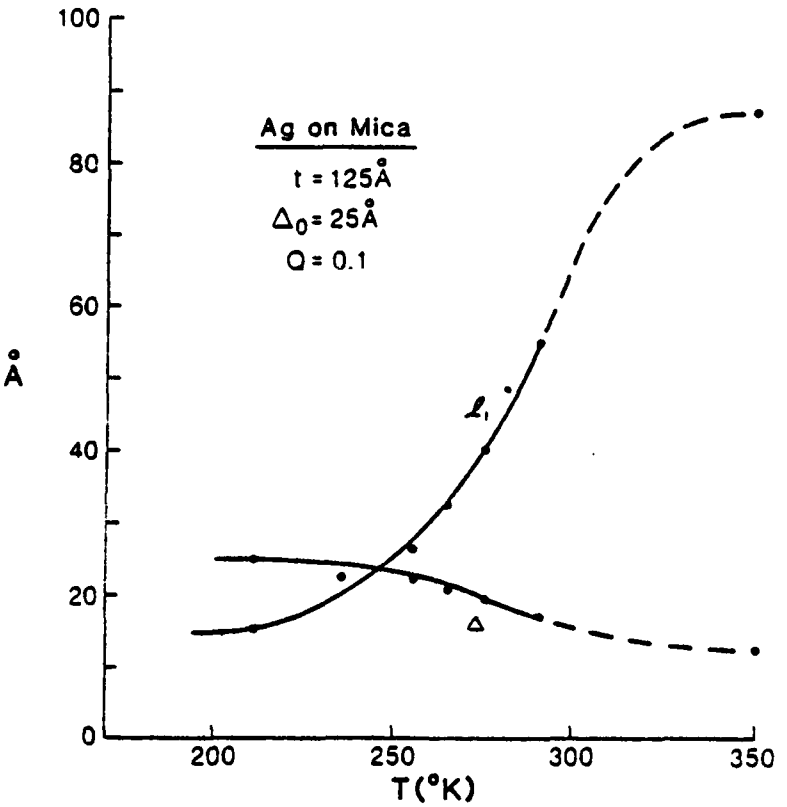


Fig. 5 Surface structure parameters Δ (surface layer thickness) and ℓ_i (surface mean free path) vs. stability temperature T_c, as deduced from Fig. 4 using a model theory of surface disorder scattering of conduction electrons.

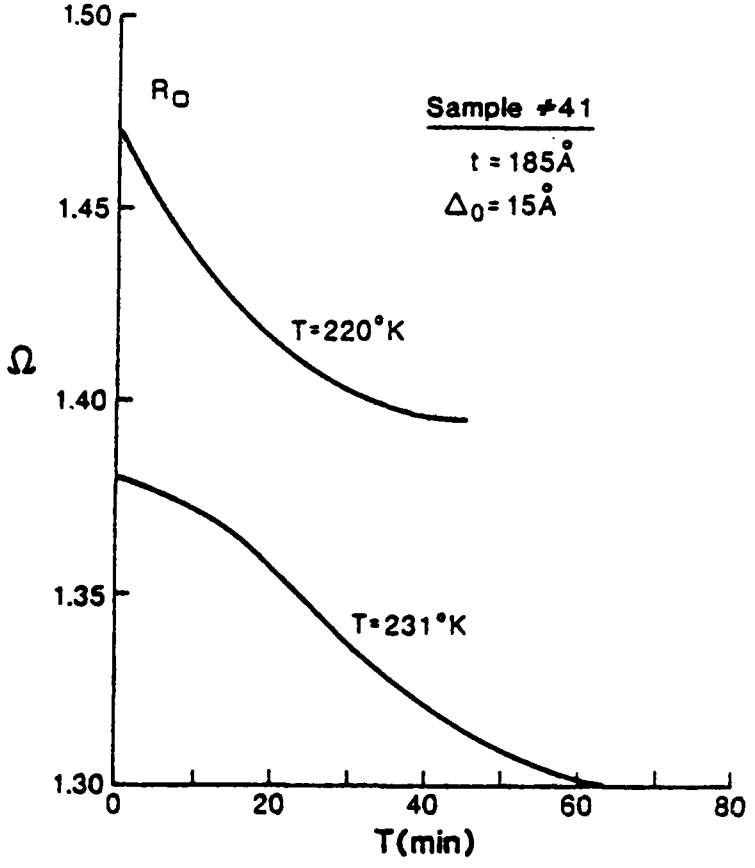


Fig. 6 Time variation of silver film resistance when its temperature is raised to $T = T_c + 10°K$, for a film of 185Å with an initial rough overlay of 15Å.

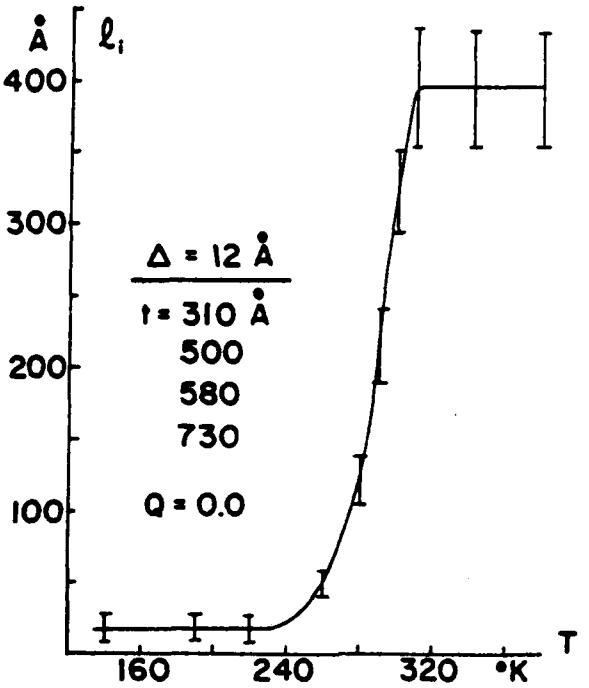


Fig. 7 Recent resolution of the data for ℓ_i previously presented in Fig. 5.

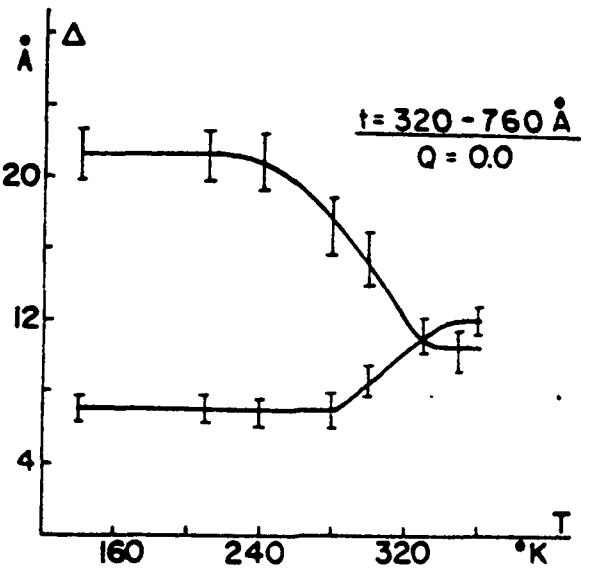


Fig. 8 Recent resolution of the data for Δ previously presented in Fig. 5.

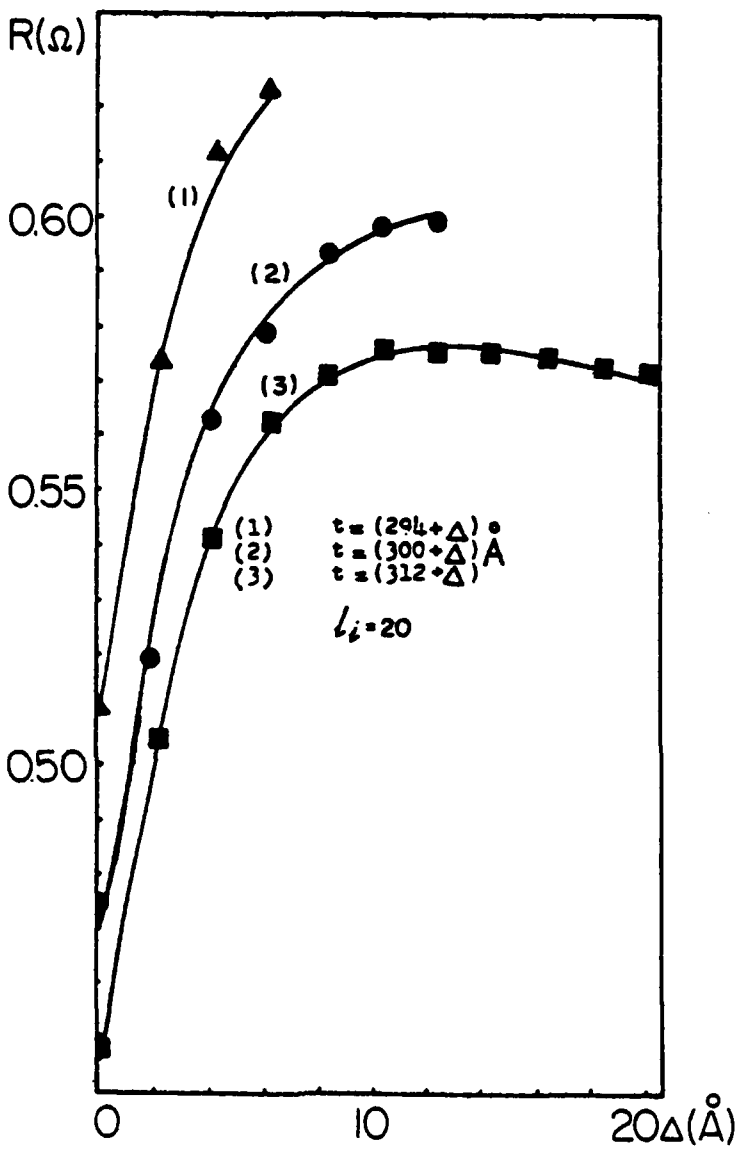


Fig. 9 Change of resistance of an epitaxial silver film of thicknesses 294 Å, 300 Å and 312 Å, when overlays of 6 Å, 15 Å, and 20 Å, respectively, are successively deposited at 140 °K. The solid curves give the theoretical fit with a common order parameter ℓ_i = 20 Å in the overlays. Between successive deposits the previous overlay is again annealed, so that the measurements begin again at lower resistance values.

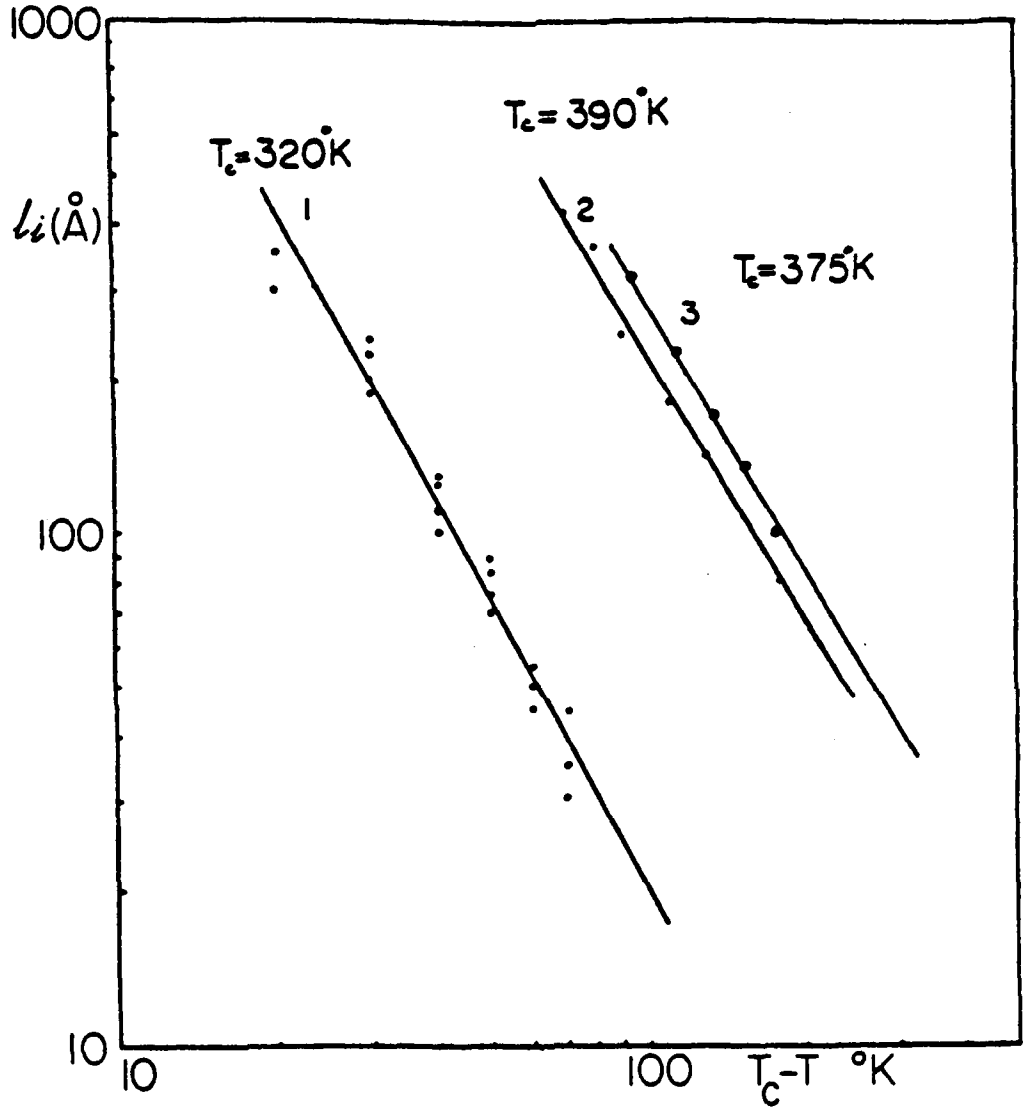


Fig. 10 Temperature dependence of the order parameter $\ell_i(T)$ in terms of a critical point formulation $\ell_i(T) = L(T_c-T)^{-n}$, for thin overlays (curve 1) and for thick films (curves 2 and 3), all deposited at low temperature.

in Surface Enhanced Raman Scattering, eds. R.K. Chang and T.E. Furtag, Plenum (1981).

4. U. Landman, R.N. Hill and M. Mostoller, "Lattice relaxation on metal surfaces, an electrostatic model," Phys. Rev. B21, 448 (1980).

5. J.P. Chauvineau, "Diffusion superficielle des adatomes observée a basse temperature par la variation de résistance électrique de films minces d'or," Surf. Sci. 93, 471 (1980).

6. Chemisorption and Reactions on Metallic Films, J.R. Anderson, ed., Academic Press (1971).

7. I.P. Grebennik, "Resistivity of polycrystalline silver film with artificial surface roughness," Phys. Met. Metall. 47, 42 (1980).

8. D.A. Glocker and M. Skove, "Field effect and magnetoresistance in small Bismuth wires," Phys. Rev. B15, 608 (1977).

9. A. Berman and H.J. Juretschke, "Size effects in resistivity of epitaxial films of silver," Phys. Rev. B11, 2893 (1975).

A. Berman and H.J. Juretschke, "Origin of the metallic field effect," Phys. Rev. B11, 2903 (1975).

10. D.J. Lischner and H.J. Juretschke, "Maxwell stresses at charged metal surfaces from thin-film elastoresistance," J. Appl. Phys. 51, 474 (1980).

11. P. Mazumdar and H.J. Juretschke, "Magnetoconductivity modulation of thin ferromagnetic films by substrate electrostriction," Phys. Rev. B19, 664 (1979).

12. P. Mazumdar and H.J. Juretschke, "Oscillatory metallic field effect and surface magnetoelasticity in thin ferromagnetic films," Phys. Rev. B19, 672 (1979).

13. H.J. Juretschke and R. Pimpinella, "Size effect in thin metallic films resulting from modified surface transport properties," submitted to Surface Science.

14. R. Pimpinella, "Metallic field effect, surface scattering and surface structure of free metal surfaces," Ph.D. Thesis, Polytechnic Institute of New York (Fall 1981).

15. J.M. Lockhart, F.C. Witteborn and W.M. Fairbank, "Evidence for a temperature-dependent surface shielding effect in Cu," Phys. Rev. Lett. 38, 1220 (1977).

16. Y. Amani and H.J. Juretschke, "Temperature dependence of the structural disorder of artificially roughened silver surfaces," presented at American Physical Society Meeting, New York (January 1983).

17. R.J. Pimpinella and H.J. Juretschke, "Metallic field effect at the free (111) silver surface as a function of temperature," presented at American Physical Society Meeting, New York (January 1983).

18. Y. Amani, "Temperature dependence of the structural disorder of artificially roughened silver surface," Ph.D. thesis, Polytechnic Institute of New York (1983).

6. DoD AND OTHER INTERACTIONS

In view of our extensive experience with the physics of strains on metal surfaces, gained largely under JSEP support on this work unit and its predecessor, we were asked by AFOSR about four or five years ago to advise it on the RADAM problem, involving the modulation of radar scattering by moving or vibrating contacts. This consultation led to a separate contract with AFOSR to determine the effects due to the intrinsic nonlinear electrical properties of weathered and rusty metal surfaces. In particular, we investigated the role of microscopic metal-oxide barriers. We found that they allow tunneling of electrons through surprisingly thick barriers, and at low energies, and that the sensitivity of such barriers to microscopic strain is large, but not quite large enough to explain the modulation effects on that basis alone.

C. X-RAY COUPLED WAVE INTERACTIONS AT CRYSTAL SURFACES

Professor B. Post and H.J. Juretschke

Unit SS3-3

1. OBJECTIVE(S)

To use the multiple interaction of x-rays in a Bragg geometry in order to develop simple and compact methods for obtaining direct phase information about the crystal scattering factors, and for characterizing the mode structure; to extend the interactions to include coupling to other waves that can modify the mode structure, and that can be used to explore non-linear interactions of x-rays; and to understand the local x-ray fields in the immediate surface region of the solid, as well as the effect of stringent boundary conditions on all modified x-ray waves, especially those originating in the interior.

2. APPROACH

High resolution x-ray diffraction in the region of n-beam interactions contains both phase and mode structure information and will also reveal effects of coupling to other waves by specific angular shifts in the diffraction directions. The experiments will be supported by theoretical studies of the predictions of n-beam dynamical theory, and by rigorous extensions of the theory to include the coupling with other waves.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

This past year has been one of consolidating various aspects of the earlier work, completing a number of experimental tasks, and putting the past year's efforts into perspective by preparing a number of manuscripts and theses.

A. n-Beam Interactions

A new formulation of the classical x-ray problem of 3-beam interactions has been developed for the purpose of better understanding the various contributions that modify the ordinary 2-beam situation as the coupling to the third beam is approached. This formulation predicts quantitatively, and in analytic form, the major features that have been observed in high resolution experiments, as well as several new effects not yet seen. A paper has been submitted to Acta Crystallographica,[26] and three talks[27-29] have been presented during this past year on this work.

B. X-ray Interactions with Thermal Phonons

The dynamical, i.e., self-consistent, theory of this problem has been completed. It predicts considerably more fine structure than has so far been seen in the best experiments, but it also reproduces well their gross features. The theoretical framework includes a generalization of ordinary dynamical theory that allows the coupling of the proper (Bloch) modes via phonons, as well as the strict observation of all appropriate boundary conditions. A talk[30] at an American Physical Society meeting was presented in January 1983.

C. Phase Determination in Mosaic Crystals

The base line structure of Bragg reflections in the neighborhood of n-beam interactions has been observed without too much difficulty in crystals of average quality (mosaic crystals), when the proper conditions of high resolution are met. This makes it clear that our method for the determination of phases of x-ray structure factors, which relies on the asymmetry of the base line structure relative to the n-beam interaction point, is in no way restricted to nearly perfect crystals. A paper on these results has been accepted for publication in Acta Crystallographica.[31] A theoretical model for correlating these features to other asymmetries in the propagation of x-rays in this region has also been worked out. A paper explaining this theoretical model has been accepted for publication in Acta Crystallographica.[32]

4. STATE OF THE ART AND PROGRESS DETAILS

With the advent of new high intensity sources covering a continuous frequency spectrum, such as the synchrotron, x-rays are assuming a significant role as a tool in surface exploration, similar to their traditional role in the structure determination of bulk solids. At the same time, the new degrees of freedom opened up by these sources now allow including x-rays as a fully accessible part of the conventional electromagnetic spectrum, with the same possibilities for application that have been obtained historically, for example, in the microwave and visible parts of the spectrum. From this point of view, however, both the conventional treatment of x-ray interactions, as well as the experimental techniques being used have to be reexamined and redirected in order to make them suitable for exploiting these new directions.

The basic interaction of x-rays with crystals, as formulated more than fifty years ago, has always included the possibility that more than the two x-ray fields of common diffraction phenomena, the so-called incident and diffracted beams, can couple within the crystal.[1] Such more general coupling, designated as the n-beam case, however, occurs only under stringent conditions of geometry and wavelength, and is observable primarily in crystals of good perfection and when using x-ray sources of high monochromaticity and high angular definition. Among the most startling consequences of these n-beam interactions have been (a) the phenomenal decrease in linear absorption of beams traveling along some special directions, allowing highly collimated x-rays to pass literally through centimeters of normally highly absorbing material,[2] and (b) the fact that n-beam interactions can be used to extract the phases of individual scattering factors,[3,4] information that has been one of the

long sought goals of crystallographers because it is needed to put the determination of molecular structure from measured diffraction intensities on a fully deterministic footing. The underlying theory of these effects, the dynamical theory, is fully formulated, and various aspects of the properties of some special 3 to 6-beam interactions have been investigated both experimentally and theoretically.[5-7] But the complexity of the theory when applied to n strongly coupled beams and the increasing diversity of possible experimental arrangements for observing the effects have, so far, prevented the design of a truly optimal and efficient experimental technique that is applicable to a wide range of crystals and n-beam combinations. At the same time, there is still no formulation of a straightforward and unambiguous method for interpreting n-beam data with respect either to the propagation characteristics of particular x-ray beams, or to phase information about a structure.

The same stringent experimental conditions needed for the appearance and study of n-beam interactions also allow detailed observation of the coupling of x-rays to other types of waves propagating within the crystal. These waves, such as for instance, phonons, photons in the visible, and polaritons, usually have energies small compared to x-ray photons. But they can contribute enough momentum to the interaction to produce small but observable angular shifts in and out of a particular n-beam interaction region. In principle, then, such other waves can steer x-rays, and can therefore perhaps control their propagation characteristics, or their absorption within the crystal host. They are, of course, also of interest in their own right in elucidating higher order couplings and nonlinear interactions in crystals. The experimental work in this area has usually been confined to first order interpretation. Most of the theoretical work has concentrated on applying the conventional methods for treating nonlinear interactions in matter without taking into account the specific nature of the x-ray fields of interest.[9] This includes the fact that x-rays are dispersive in the configurations of most interest when interacting with other types of waves. For example, one usually ignores that x-ray fields form coupled plane wave modes even when the conditions for diffraction are not fully satisfied, so that the initial and final states of the x-ray in a scattering process are not plane waves, but rather Bloch waves, composed of two or more coherent plane waves of fixed relative amplitudes and propagation vectors. Obviously, since the effects of such scattering are only observable outside the crystal, the outside plane wave modes must also relate properly to the inside coupled modes through rather intricate boundary conditions before reliable predictions of observable phenomena can be made.

Since both the interactions of x-rays with each other, and the interactions of x-rays with other types of waves, call for essentially the same theoretical formalism, and for similar conditions for experimental observation, they together form a natural combination in the research project that we propose here. The work will concentrate on such interactions in the neighborhood of the surface of entry of the incident x-ray beam, to be observed by x-rays emerging from the same surface. This is the so-called Bragg case, and it has been chosen partly because the experimental observations are more sensitive for reflected beams than for beams transmitted through a crystal slab (the Laue case), and partly because the mode boundary conditions only involve one interface,

for a sufficiently thick crystal. A careful experimental investigation under such an arrangement requires, of course, not only that the crystal structure be very good on the whole, but also that the surface region of interest have comparable regularity, so that homogeneity persists throughout most of the region of interaction.

The work that we propose will have two major specific goals: (a) to develop simple and straightforward procedures for determining the phases of scattering factors in complex crystals and of less than perfect order; (b) to establish both experimental and sound theoretical criteria for the systematic study of the interaction of x-rays with various other kinds of waves. As we will see below, these two goals will, in fact, involve many other aspects of coupled wave interactions, including those already mentioned in the introduction. Part of the work will consist of an extension of research already begun, and the remaining portion will deal with related but new aspects whose importance has emerged as a consequence of seeing connections of our own work with current efforts in neighboring areas of research at other laboratories.

In order to understand the results of our most recent work, it is necessary to place them in perspective by including some immediately preceding stages of progress.

Since the concerns in the first two main areas, n-beam interactions and coupling with other waves, involve different levels of approach, the first immediately drawing on a full-fledged theory while the second began with very little sound theoretical foundation, the two areas will be discussed in sequence. It will soon become clear, though, that both are intimately interconnected.

A. n-Beam Interactions

Largely for historical reasons, because the first indication of anomalous interaction features (Borrmann effect) had been found in the propagation through crystals, our work concentrated on experimental consequences in the Laue case, which detects the diffracted beams moving in the forward direction relative to the incident beam, beyond the exit surface of the crystal. This configuration was also employed in the seminal paper on phase determination in 3-beam interactions by Professor Post[3] that formed the basis of much of our subsequent work. A number of specific 3-, 4-, and 6-beam Laue interactions have been studied in detail,[5,6] partly in order to understand the development of their mode structure as one passes through the n-beam region along specific directions, and partly in order to try to abstract those common characteristics that should allow for at least a qualitative prediction as to what happens in other cases. Most of the studies employed computer experimentation because some of the constraints for observing effects of coupling were not known initially -- and subsequently turned out to be beyond those of x-ray sources other than the synchrotron, so that actual experimentation would have been inconclusive. The major results which emerged from these studies were:

(1) The mode structure is an extremely sensitive function of location within the n-beam region, with relative amplitudes, propagation vectors (complex), and polarizations of the various plane wave com-

ponents varying very rapidly with changes of seconds of arc.[6,10]

(2) For any specific n-beam case, the polarization of the incident beam is an important variable for optimizing the amplitudes of the least absorbing components of the n-beam mode, and therefore also optimizing the transmitted intensity.[6]

(3) The n-beam interaction can be used to produce highly collimated pencils of x-rays (divergence of a second of arc or less) with specified polarization. This suggests it as a highly selective optical component for x-ray experiments at synchrotron facilities.

(4) In contrast to the complications of the mode structure, the dispersion surface in the n-beam region shows many regular features, mostly dictated by geometry and continuity, which allow both an immediate qualitative identification of the predominant mode components, as well as an identification of the invariant phase factors associated with the particular interaction.

(5) Applying the ideas of (4), phase information can be obtained by examining the asymmetry in the intensity of the transmitted beams as one passes through the n-beam region.[11]

(6) While the procedure of (5) has been successful for nearly perfect crystals, such as Ge and Si, our experimental work on less perfect crystals, especially organic ones, has not given unambiguous results. From this work, however, it seems plausible that if the same investigation is carried out at longer x-ray wavelengths such as supplied by synchrotrons, phase effects should become more pronounced.

(7) Our most recent preliminary results have extended the notions based on (4) to the Bragg case, again looking for the left-right asymmetry in diffracted asymmetries as one passes through n-beam regions along a weakly reflecting 2-beam line. These results are extremely promising. They show pronounced and systematic asymmetries under only moderately stringent experimental conditions, for practically all the n-beam interactions crossed by the two beam line, with the asymmetry in the simple cases always in the direction predicted by inspection of the dispersion surface. Because of its ease, this method may be the practical one that would make phase determination routinely possible.

Our results this past year on n-beam interactions are the following.

The theoretical approach developed in the last part of the previous (FY82) period and in late 1982, proposing a simple method for determining the phase of x-ray structure factors (Refs. 24, 25), has been generalized as part of a new systematic formulation of 3-beam interactions. This formulation relies on rewriting a 3-beam interaction in a form using the usual 2-beam variables as primary variables, with all 3-beam effects occurring as modifications of these variables, and of the 2-beam structure factor, and consequently of the mode structure. We have shown that such a formulation of the 3-beam case is possible, and that, in particular, it lends itself to describe and analyze the onset of the influence of the third reciprocal lattice point on x-ray fields and

intensities in compact analytic form, that can also be interpreted in terms of simple physics. This is in contrast to all earlier attempts to handle the 3-beam case, either by a computer analysis, which, though correct, reveals very little physical content that can be applied in general, or by analytic formulations that start out at the 3-beam point, and are usually so complicated that again the physics remains obscure, and the applicability to a variety of physically interesting cases is not simple. In the range of the onset of 3-beam effects, the influence of the third reciprocal lattice point is parametric, rather than dynamic, so that the problem reduces to discussing a modified 2-beam case.

In lowest order of modification, this new 2-beam case has the following properties:

1) It describes the typical 2-beam behavior, but around a shifted origin in reciprocal space.

2) 2-beam scattering is controlled by a modified structure factor, that is strongly asymmetric with respect to the direction of approach of the 3-beam point, in fact going close to zero on one side.

3) The onset of all 3-beam effects occurs on a universal scale x that takes into account the pertinent angular factors and the coupling strength of any particular 3-beam interaction, and both conclusions 1) and 2) above take a very simple form with respect to x for all 3-beam interactions. For example, the modified structure factor (for σ polarization) becomes $F^{\sigma}_{HL} = F_H(1+1/x)$, so that it vanishes when $x = -1$. This value of x may correspond to fractions of a minute of arc, or also to many minutes, depending on the 3-beam interaction in question. Therefore the formulation contains automatically very long range effects.

4) When $|x| \sim 1$, the usual σ and π plane polarized modes start to couple, so that the proper modes of propagation within the crystal are actually elliptically polarized.

5) The effect of a sequence of neighboring 3-beam points is additive in the structure factors, in this lowest approximation, and results in a diffracted integrated intensity profile as a function of angle showing considerable structure. Figure 3 gives the computed result, based on our theory, of a sequence of Ge 222 3-beam interactions. The curve is based entirely on information of the geometry and of the structure factors of the separate 2-beam cases. Figure 3 should be compared with the experimental results taken from the high resolution experiment of References 18 and 22 and reproduced in Fig. 4. The agreement between the two figures is highly satisfactory, especially considering that the theory does not include any of the broadening and divergences that are unavoidable even in a high resolution experiment.

6) Higher order interactions, such as for 4 or 6 beams, are

formally described as two or four coincident 3-beam interactions, and the rules for determining the diffracted intensities, and especially their asymmetry, follow from the weighted contribution of all 3-beam interactions to the compound modified structure factor. Effects following this particular rule have been observed in the recent experimental part of our program (Ref. 11 and 21), although some of the detailed predictions with respect to the proper weighting remain to be verified.

7) The physics underlying these effects follows from the distortion the dispersion surface caused by 3-beam coupling: the dispersion surface sheets are displaced in common, as well as relative to each other. The relative displacement, in particular, affects the diffracted intensity, and when this intensity vanishes, the dispersion surface sheets cross. This behavior is completely analogous to the crossing of bands in the electron theory of solids, and many of the consequences are the same in both cases.

8) The first order theory forms the natural starting point for carrying the solution to the next order, which involves both the elliptical polarization of the normal modes, and the probable removal of some of the degeneracies implied in the band crossings of the first order solutions. This removal will affect the actual diffracted intensities near their minima, but without altering the overall asymmetry predicted by the first order solution.

The detailed exposition and discussion of the above effects have been incorporated in a paper submitted for publication in Acta Crystallographica.[26]

B. Interactions of X-rays With Other Waves

As already mentioned with respect to this topic, one of the major thrusts during the current program was directed towards extending the conventional dynamical theory of x-rays in crystals to systematically include the coupling with other waves within this framework. We immediately chose the Bragg case for x-rays, for reasons like those already mentioned, and confined ourselves entirely to the two-beam case in order to learn how to handle the desired coupling in the least non-trivial dynamical setting. As with all such theoretical probing tied to a simple physical model, we developed a number of physically motivated approaches to the problem before formulating a general theory. In order to give the theory concrete content, we focused on furnishing a quantitative interpretation of two experiments in the recent literature. In one experiment, satellite diffractions were observed when intense beams of phonons travel parallel to the reflecting surface.[12] In the other, details of the fine structure of x-ray-thermal phonon scattering were displayed in an arrangement in which the incident beam entered away from the Bragg angle and the coupling produced output in a whole range of directions.[8] A third focus of the theory centered on predicting the consequence of the interaction between x-rays and lasers in ionic crystals via the polariton modes, and used this prediction to set up the appropriate experiment. Progress in all of these directions

has been very good. However, rather than recount the sequential development through the course of the program, the summary below will list the current state of this investigation. Much of the work is nearing completion at this point. Because this particular research started off in a new area from the beginning, there have been no publications so far, but we expect that before the actual start of the new proposed work several papers will be ready. Much of the work is related to a Ph.D. thesis completed in 1982.[13]

(1) We have found a fully analytic solution of the six-beam case of an x-ray interacting dynamically with monochromatic high intensity phonons moving parallel to the crystal surface. In this formulation the problem factors into cubic and quadratic equations, and gives explicit solutions for the three-sheet dispersion surface and the sixtuplets of (Fig. 1) x-ray field amplitudes associated with each point on this surface. We can therefore exactly solve the boundary condition problem, and determine explicitly all external reflected fields for arbitrary intensity of the phonon beam. This exact solution has been extremely useful in exploring the effect of changing the various parameters in the problem, and especially under what conditions the usual weak coupling approximations hold. We have studied both model problems and the specific interaction of Mo radiation reflected off the InSb (004) planes. Among the startling consequences of the analysis is that not only does this interaction lead to slightly frequency shifted satellites of the main reflection, but this reflection itself also shows satellite structure, coming off at different angles (Fig. 2). To our knowledge this effect has never been reported, and it suggests that future investigations in this field should include a much more careful analysis of the reflected beams than usual. This work has now been published.[14]

(2) We have formulated the dynamic n-phonon problem, and set up the explicit equations for the $2n + 1$ sheeted dispersion surface and the associated fields. Introduction of the boundary conditions leads to a $2n$ by $2n$ determinant for the relative incident fields, and the corresponding expressions for the reflected fields. When the coupling is weak, these expressions reduce to explicit modifications of the uncoupled solutions for all but the main reflection, which has still to be solved for a modified dispersion relation. Our expressions so far show that this modification contains parts equivalent to the Debye-Waller factor but also other contributions that only appear when the full problem is considered. In particular, the theory also predicts a decrease in peak and a narrowing of the main beam, with satellite structure in the wings.

(3) We have evaluated the modification of the main reflection intensity by many thermal phonons, when the reflection occurs away from the Bragg angle, and find that the multiphonon coupling leads to an appreciable reduction below the intensity of the normal 2-beam case. Numerical calculation is in progress to assess this reduction quantitatively, and an accurate value may, in fact, have to include fine features of the dispersion surface and of the phonon distribution in order to avoid a possible divergence.[15]

(4) The interaction of x-rays with laser-induced polaritons has been analyzed from the point of view of energy transfer and it is found

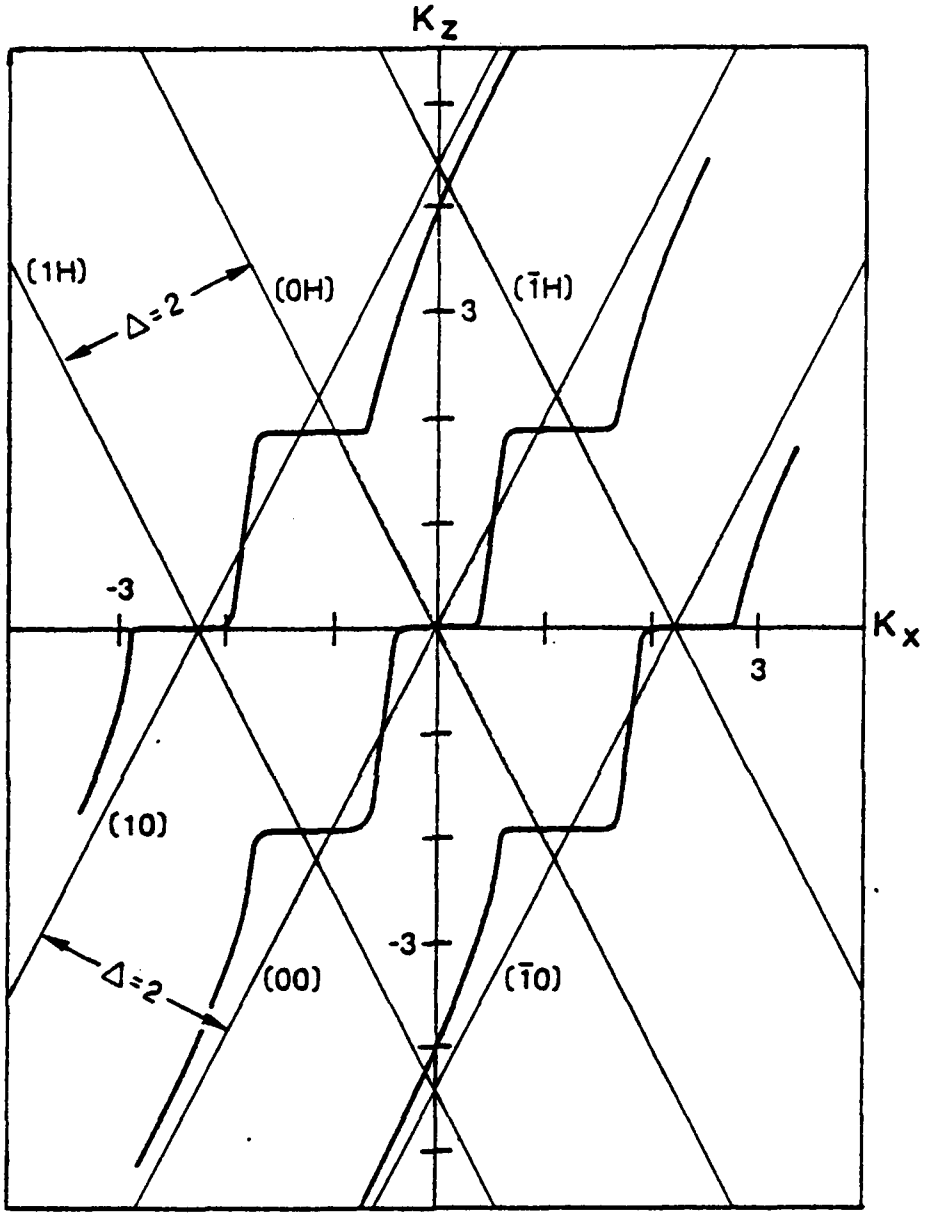


Fig. 1 3-sheeted dispersion surface in the K_x-K_z plane for a 2-beam symmetric Bragg x-ray mode interacting with a phonon of q = 2.5 parallel to the crystal surface. The slanted light lines are the asymptotic surfaces of the Ewald spheres, and the heavy horizontal portions of the dispersion sheets give the regions of strong diffraction.

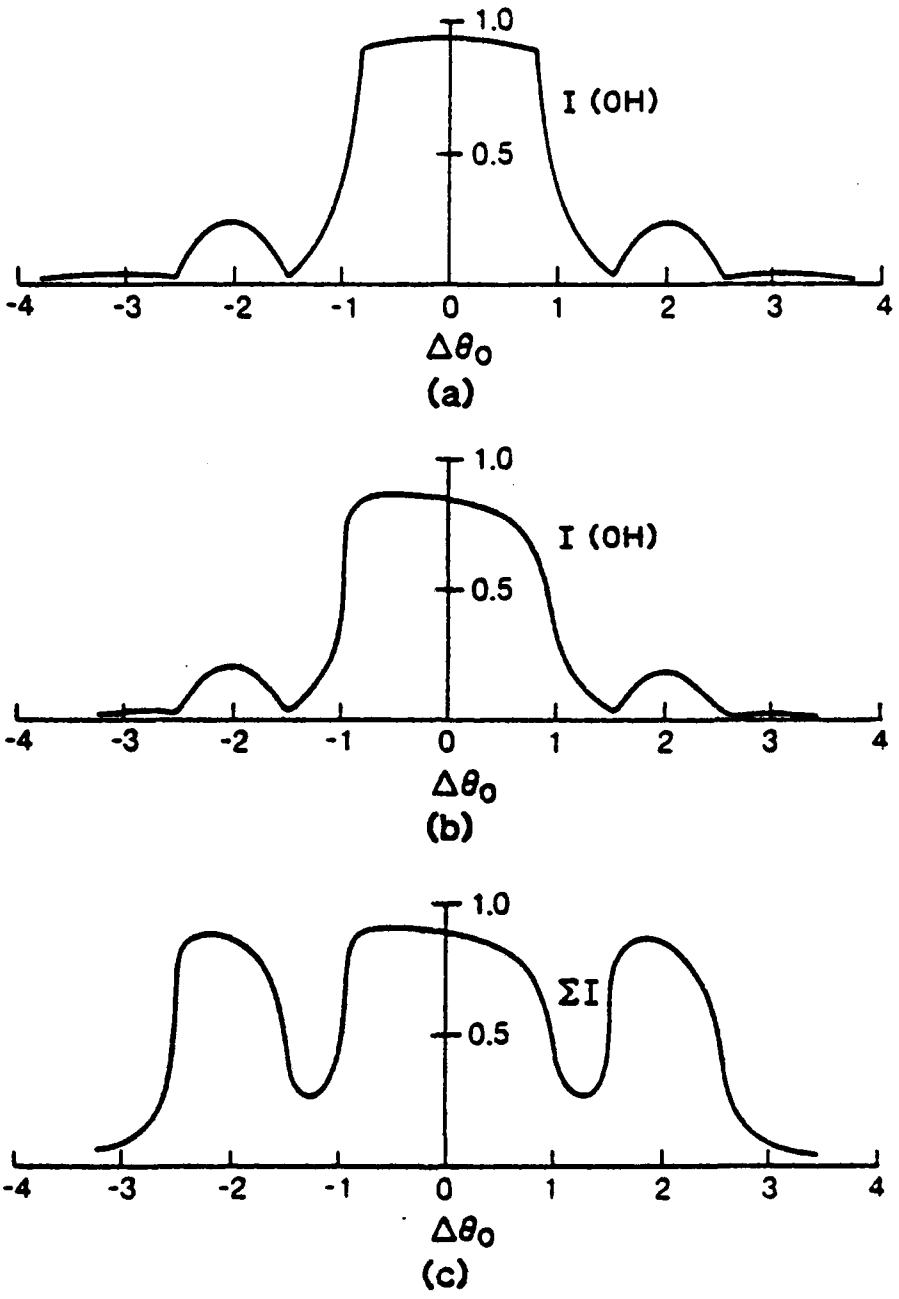


Fig. 2 Diffracted intensity I(OH) of the unmodified x-ray field in the presence of an intense phonon beam as in Fig. 1. (a) No absorption; (b) 5% absorption; (c) Total diffracted intensity for case (b) if the output detector integrates over all angles of emission, as a function of incident angle $\Delta\theta_i$.

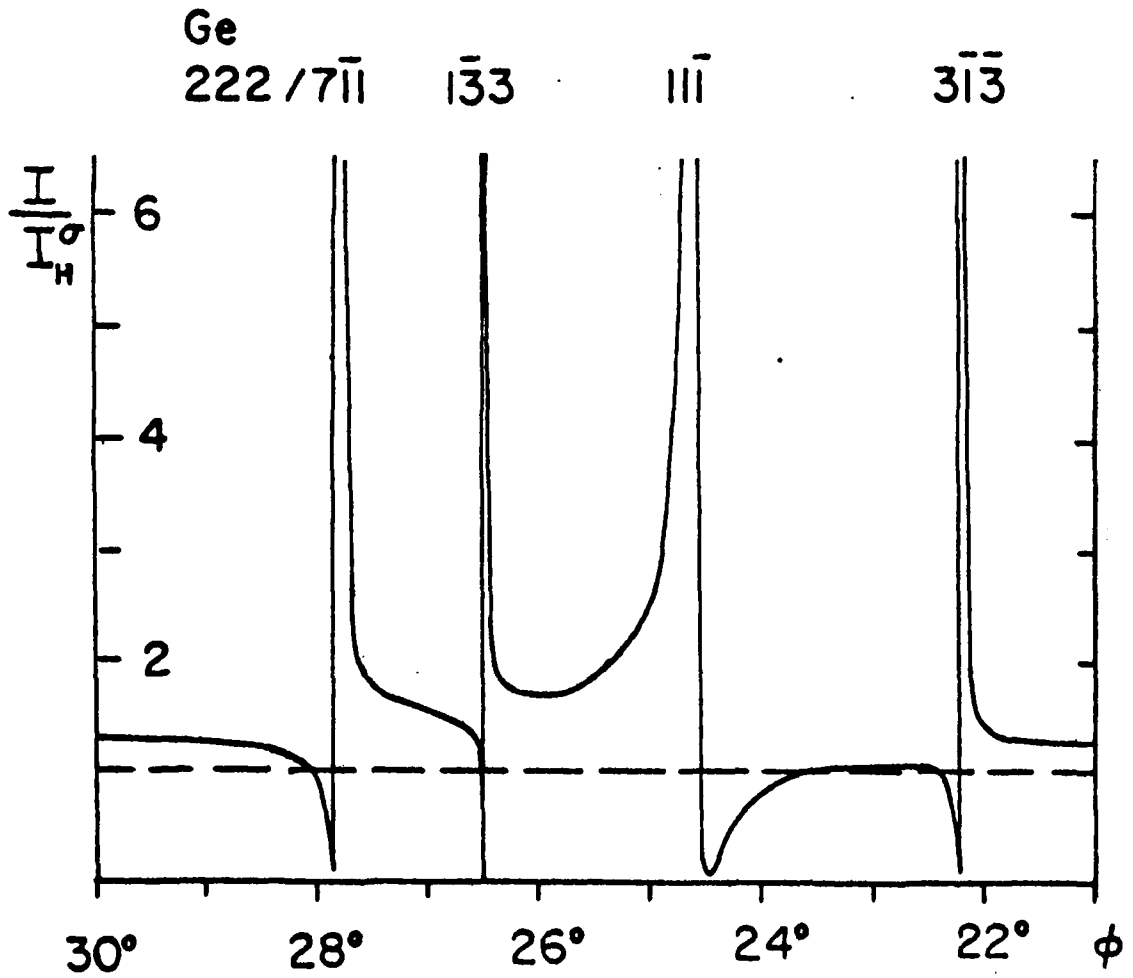


Fig. 3 Theoretical prediction of the integrated intensity of the Ge 222 reflection as a function of the azimuthal angle ϕ. Peaks occur whenever another set of crystal planes is in simultaneous diffracting position.

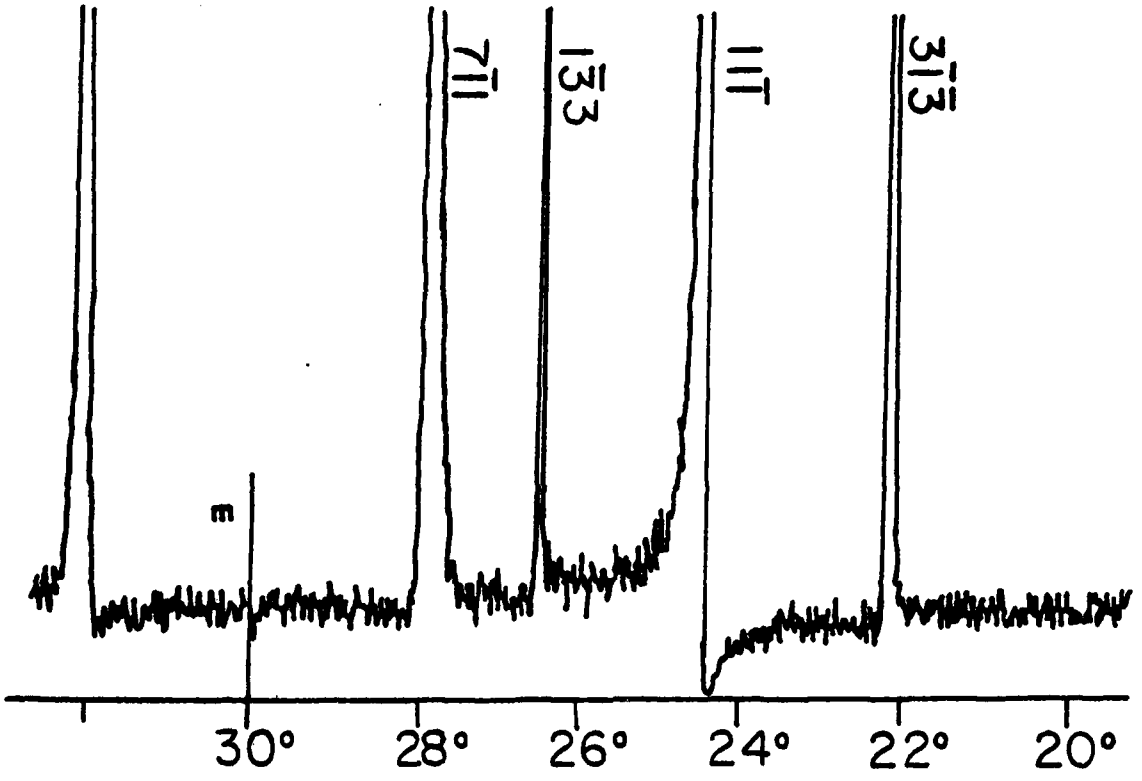


Fig. 4 Experimental chart of the same Ge 222 reflection discussed in Fig. 3, for incident beams of 40 seconds of arc divergence.

that under optimal conditions, using normal laser intensities, the weak coupling approximation for the phonon part of the polariton is justified. This has made it possible to adapt directly the formalism worked out under (1), above, to single phonons traveling in other directions, but with the additional condition that energy transfer between polariton and x-ray is not completely negligible. This implies that the coupled modes contain fields of differing frequencies that refer to different dispersion surfaces. However, our formalism lends itself readily to incorporate this complication. One immediate consequence is that the expected 6-beam coupling reduces to 4-beam cases, i.e., polaritons are predominantly either absorbed or emitted, but not both. We are in the process of determining the optimum experimental configuration for observing laser induced switching of x-rays in a selected material, such as MgO, with the actual experiment being one of the tasks of the proposed research work.

The results obtained this past year related to x-ray interactions with thermal phonons are the following.

The dynamical solutions of the coupling of diffracting x-rays with phonons that we have developed confirm many of the usually encountered features that follow from purely kinematic considerations. In addition, they also contain new aspects that emerge only when the exact nature of the x-ray modes resulting from the distortion of the dispersion surface by the coupling is taken into account. Furthermore, the actual excitation of the interaction also depends on the boundary conditions that have to be satisfied for both the incident and the scattered x-ray modes.

We have completed the calculation of this interaction for the particular case for which high resolution experimental data exist in the literature (Ref. 8), and have also included the effects of broadening of the incident beam and of the method of detection, which here, because long wavelength phonons produce only very small angular shifts in scattering, can substantially influence the measured effects. Typical results are shown in Fig. 5, where the top graph reproduces the experimental results, and the two lower graphs show the theoretical expectations for the broad low peaks of the top graph. (The sharp peak there represents a residual specular contribution). As seen in the graphs of the two theoretical curves, there is general agreement with experiment about the location and the overall angular spread of the interaction. In addition, theory shows considerable fine structure as a function of the angle of detection θ_{out} that is not visible in the experimental results. Much of this structure is probably hidden in the noise level of the experiment. If the dashed lines are taken as an estimate of such a level, only the features above it should be visible, and this improves the agreement. Of course, at the time the experiments were done there was no detailed theory and therefore little effort was made to look for such fine structure. Cooperative attempts are now under way to repeat some of the experiments now that it has become clear what one must look for, as well as what complications (e.g., n-beam interactions) one must avoid.

In order to bring the theoretical results to a fully absolute scale, a further averaging over the energy spread of the incident x-rays, and

over any long range mosaic character of the crystal, is required. Since there exists no independent information on the properties of the crystals used in the original investigation, any such corrections can only be tentative at this time.

C. Phase Determination in Mosaic Crystals

A crucial question about the practicality of observing phase information by the methods we are pursuing is the extent to which the necessary effects persist when dealing with imperfect crystals, because very few crystals can be grown to the requisite purity and perfection so that the full dynamical theory of x-ray diffraction applies. Various corrections have to be made that, in a number of instances, such as the absolute total diffracted intensity, lead to qualitatively different responses in perfect and imperfect (mosaic) crystals.

To obtain answers to this question we have carried out a primarily experimental program that has tested out the appearance of the fine structure in Renninger-type experiments on a number of centrosymmetric crystals of average quality. When the emphasis is focused on minimizing the divergence of the incident beam in all directions, on monochromaticity, and on studying the details around the base line near the n-beam interacton rather than at the full interaction point itself, our work shows that there is sufficiently clear-cut information in the intensity asymmetries to allow a determination of invariant phases in most interactions, at least in centrosymmetric crystals.

As an example, our results on $ZnWO_4$, analyzing 53 interactions, of which 40 represent 4-beam couplings, agree fully with the phases of the known structure. An example of the fine structure obtainable under the appropriate conditions is shown in Fig. 6. These results have been prepared for publication, and have been accepted for publication in Acta Crystallographica.[31]

The immediate basis for identifying phases in such crystals has been the correlation between the relative orientation of the maxima and minima in the integrated intensity on both sides of the interaction point, with changes in the absorption coefficient of the most strongly excited sheet of the dispersion surface in the same regions. An explanation of this approach, and its justification in terms of a qualitative interpretation of pertinent features of the dispersion surface in the vicinity of the n-beam interaction point, has been accepted for publication in Acta Crystallographica.[32]

D. X-ray Interactions with Lasers

Among the other results, the analysis of x-ray interaction with lasers via polaritons has been completed, and a computer search is under way for optimizing the observation of this interaction relative to crystal host, to laser wavelength, and to angular shift of the laser- controlled beam. Incidentally, one of the results that has emerged from the study of x-ray phonon interactions, namely the existence of new selection rules suppressing the role of phonons within or close to an x-ray reflection, also affects this interaction, and eliminates the most obvious configuration of x-rays and lasers in producing coupling. The search must therefore

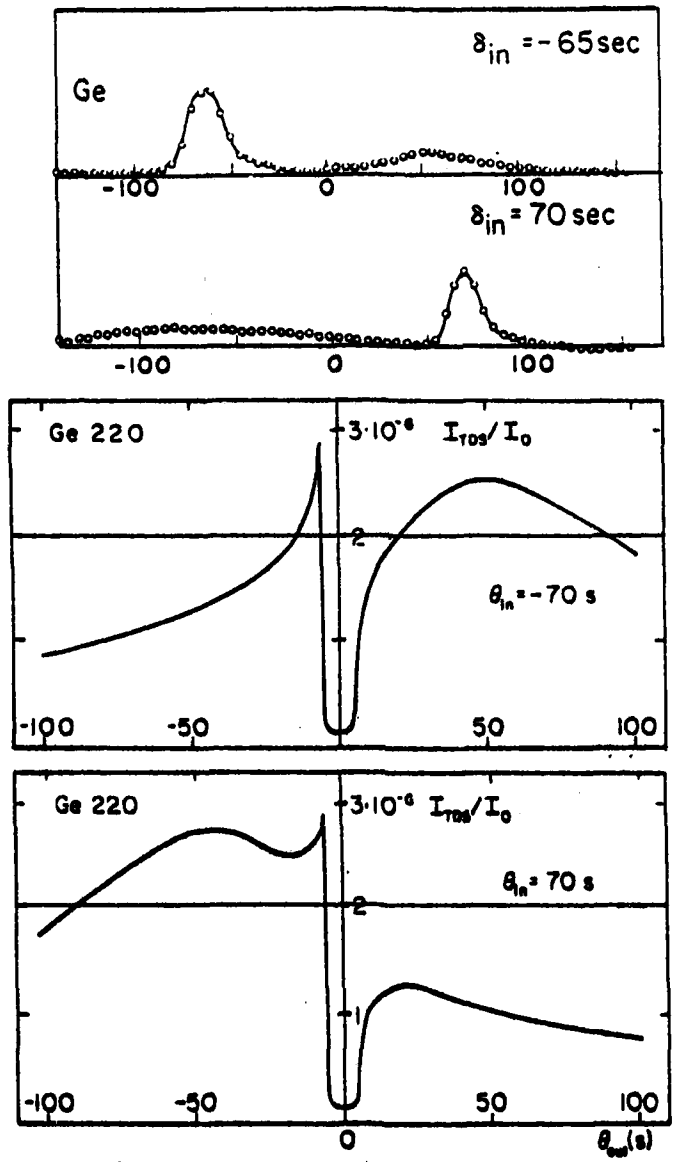


Fig. 5 Thermal phonon-modified x-ray diffraction for Ge 220. a) Experimental data for incident x-rays -65 and +70 seconds away from the Bragg angle. b) and c): Dynamical theory of this effect for ±70 seconds of the incident x-ray angle, as a function of the output angle θ_{out}. The horizontal dashed line is an estimate of the background level.

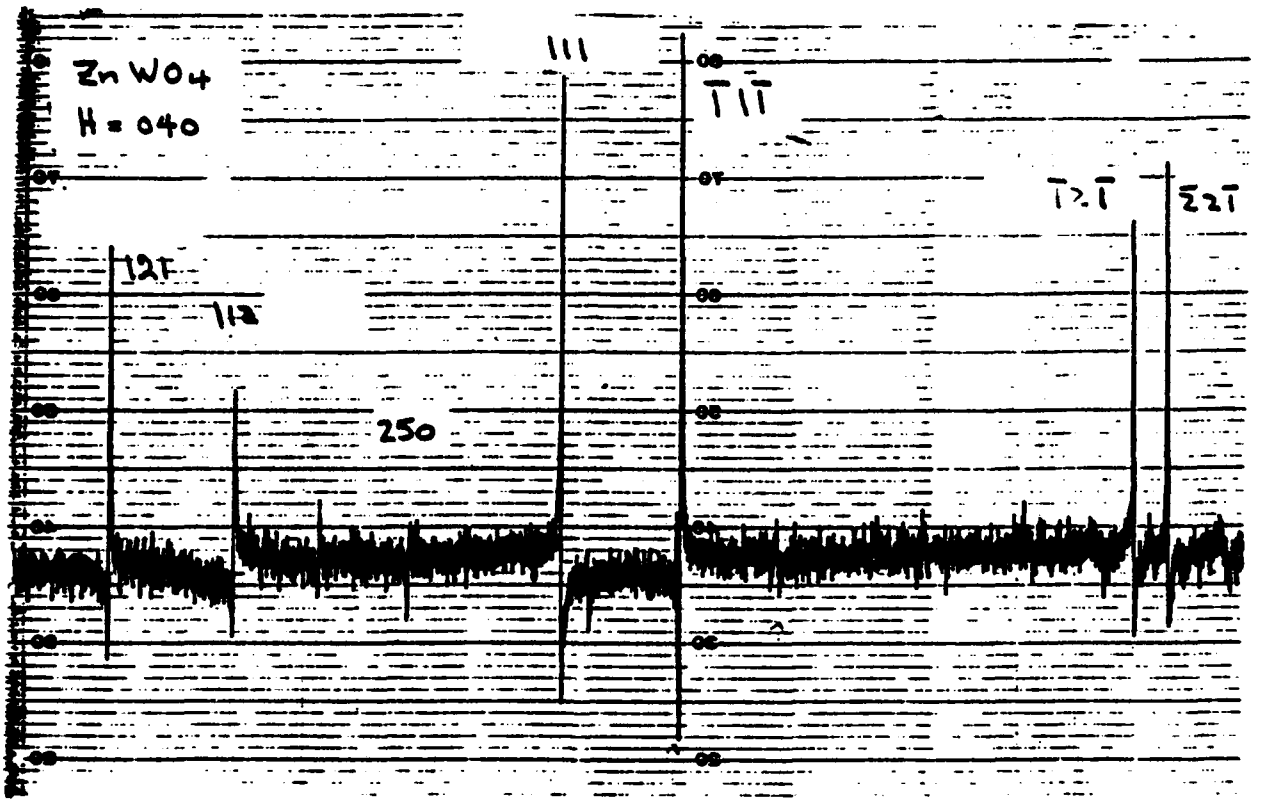


Fig. 6 Experimental chart of the same type as in Fig. 4, but for the 040 reflected intensity of $ZnWO_4$, a crystal of average crystallographic quality (mosaic crystal).

concentrate on the wings of the x-ray line, and especially on that side that is characterized by a small absorption coefficient.

5. REFERENCES AND PUBLICATIONS

1. B. Batterman and H. Cole, Rev. Mod. Phys. 36, 681 (1964)

2. G. Borrmann, Trends in Atomic Physics, ed. O.R. Frisch, Interscience, N.Y. 1959.

3. B. Post, Phys. Rev. Lett. 39, 760 (1977).

4. B. Post, Acta Cryst. A 35, 17 (1979).

5. B. Post, S.L. Chang, T.C. Huang, Acta. Cryst. A 33, 90 (1977).

6. T. Hom, "Three Beam Dynamical Interactions," Ph.D. Thesis, Polytechnic Institute of New York (1979).

7. A. D. Fofanov, A. V. Kuznetsov, V. G. Razgulyaev, Sov. Phys. Cryst. 21 15 (1976).

   A. M. Afanasev and V. G. Kohn, Acta. Cryst. A 32, 308 (1976).

   M. Umeno, Phys. Stat. Sol. (a) 37, 561 (1976).

8. P. Eisenberger, N. G. Alexandropoulos and P. M. Platzman, Phys. Rev. Lett. 28, 1519 (1972).

9. P. Eisenberger and S. L. McCall, Phys. Rev. Lett. 26 684 (1971).

10. P. Wang, "Direct Experimental Detection of X-ray Phases from Intensity Measurements" Ph.D. Thesis, Polytechnic Institute of New York, (1979).

11. P. Gong, "An Experimental Determination of the Phases of Mosaic Centrosymmetric Crystals," Ph.D. Thesis (January 1983).

12. S. D. LaRoux, R. Colella and R. Bray, Phys. Rev. Lett. 35, 230 (1975).

13. F. Robbins, "X-ray - Phonon Interaction as a Dynamical N-Beam Problem," Ph.D. Thesis, Polytechnic Intitute of New York, (June 1982).

14. H.J. Juretschke and F. Wasserstein-Robbins, "The Role of X-ray Boundary Conditions and Other Effects in Strong Dynamical X-ray Phonon Interactions," Phys. Rev. B26, 4262 (1982).

15. F. Robbins and H. J. Juretschke, "Dynamic Effects in X-ray Thermal Phonon Interaction in Symmetric Bragg Reflection," to be published

16. P. L. Cowan, J. A. Golovchenko and M. F. Robbins, Phys. Rev. Lett. 44, 1680 (1980).

17. M. Laue, Röntgenstrahl - Interferenzen Akademischer Verlag, Frankfurt (1960).

18. J. Nicolosi, "Experimental Procedures for Determining the Invariant-Triplet Phases of X-ray Reflections," Ph.D. Thesis (June 1982).

19. B. Post, "The Experimental Determination of X-ray Reflection Phases," Warren Award Lecture, ACA Meeting, Gaithersburg, Md. (March 1982).

20. B. Post, "Practical Aspects and Implications of Experimental Phase Determination Procedures," ACA Meeting (see above), paper J1.

21. P. Gong and B. Post, "Experimental Phase Determinations," ACA Meeting (see above), paper M6.

22. J. Nicolosi, "Application of n-Beam Diffraction for Determining the Phases of Certain Reflections," ACA Meeting (see above), paper M7.

23. B. Post, P. Wang and T. Hom, "Intensity Attenuation and Enhancement in 3-beam Diffraction," Z. Naturforsch. 37a, 528 (1982).

24. H. J. Juretschke, "Invariant-Phase Information of X-ray Structure Factors in the Two-Beam Bragg Intensity Near a Three-Beam Point," Phys. Rev. Lett. 48, 1487 (1982).

25. H. J. Juretschke, "Non-centrosymmetric Effects in the Integrated Two-Beam Bragg Intensity Near a Three-Beam Point," Phys. Letters 92A, 183 (1982).

26. H.J. Juretschke, "Modified Two-Beam Description of X-ray Fields and Intensities Near a Three-Beam Diffraction Point," submitted to Acta Crystallographica.

27. H.J. Juretschke, "Complete Crystal Structure Determination by X-rays: A Challenge," Invited talk at the City College of the City University of New York, Solid State Seminar, New York (September 22, 1982).

28. H.J. Juretschke, "A Simple Analytical Approach for Identifying Invariant Phases of Structure Factors in Multiple Beam X-ray Diffraction," Invited talk at IBM T.J. Watson Research Center, Yorktown Heights, New York (January 18, 1983).

29. H.J. Juretschke, "Invariant-Phase Information of Noncentrosymmetric X-ray Structure Factors in the Integrated Two-Beam Bragg Intensity Near a Three-Beam Point," presented at American Physical Society Meeting, New York (January 24-27, 1983).

30. F. Wasserstein-Robbins and H.J. Juretschke, "Dynamical Theory of X-ray Brillouin Scattering," presented at American Physical Society Meeting, New York (January 24-27, 1983).

31. P.P. Gong and B. Post, "The Experimental Determination of Phases of Reflections from Mosaic Crystals: $ZnWO_4$," accepted for publication in Acta Crystallographica.

32. B. Post, "The Experimental Determination of the Phases of X-ray Reflections," accepted for publication in Acta Crystallographica.

6. DoD AND OTHER INTERACTIONS

(a) The American Crystallographic Association held a special session honoring Prof. Post's contributions to crystallography at the Spring 1982 meeting at Gaithersburg, Md. Both Professor Juretschke and Professor Post presented invited papers there.

(b) Some of the above proposed work involves specific collaboration with laboratories in Brazil (Prof. Chang), Greece (Prof. Alexandropoulos), and Australia (Prof. Wagenfeld). The collaboration is being proposed by them in order to participate in our research.

D. SINGLE-LAYER AND MULTILAYER THIN FILMS: THEIR FABRICATION PROCESSES AND THEIR ELECTRONIC, ACOUSTIC, AND OPTICAL PROPERTIES

Professor W.C. Wang and Dr. S. Onishi

Unit SS3-4

1. OBJECTIVE(S)

During the past few years we have developed a novel, superior and cost-effective sputtering facility, around which our proposed program is based. With that new fabrication facility, we have successfully produced certain thin films of excellent quality, and our future program builds on that expertise. We have also begun certain analyses and measurements of basic acoustoelectric effects, and we propose to continue such studies as well.

The objectives behind our proposed program fall into two broad, but related, categories.

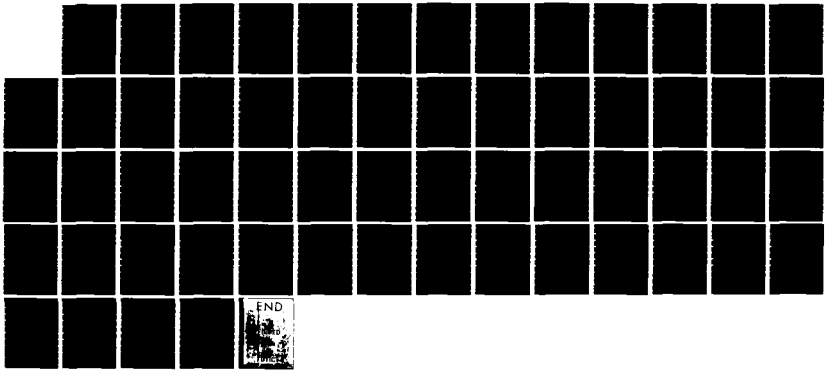
(i) The larger category involves the development of thin films of the highest quality for acoustoelectric (AE) and acousto-optic (AO) applications. In the selection of film-substrate combinations we will be guided by their potential utility in future monolithic AE or AO devices, although this thin film research should be of benefit well beyond this direct interest. The investigations in this category require two fundamental aspects. The first is that, in the course of developing films of the highest quality, we will need to better understand and to improve the fabrication processes themselves, which are basic to all thin film deposition. The second aspect is that we must perform various analyses and measurements on the electronic, acoustic and optical properties of these thin films, and on the propagation characteristics of the thin film structures that utilize them.

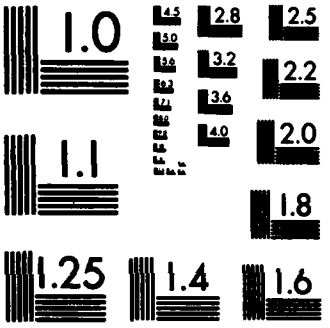
(ii) The related broad category involves basic acoustoelectric processes relating to monolithic SAW devices, and is of interest in itself but is relevant here because the proper understanding of these processes affects the measurement and evaluation of the properties of the thin films under study, and may even influence the selection of optimum film-substrate combinations. The objective here is to analyze various acoustoelectric interactions in thin film structures taking into account the surface and interfacial conditions, and then to experimentally verify the validity of these analyses.

2. APPROACH

Films deposited on various substrates by different means (including our newly developed sputtering system) and under different conditions will be systematically investigated, so that the optimum deposition conditions, such as temperature, pressure, deposition rate

AD-A136 290 BASIC RESEARCH IN ELECTRONICS (JSEP) JOINT SERVICES 3/3
ELECTRONICS PROGRAM..(U) POLYTECHNIC INST OF NEW YORK
BROOKLYN MICROWAVE RESEARCH INST.. A A OLINER
UNCLASSIFIED 30 SEP 83 POLY-MRI-1432-83 AFOSR-TR-83-1132 F/G 9/5 NL





and substrate preparation for a certain film on a particular substrate, can be found. Such information will then be translated for application to our own sputtering system, where appropriate. Both conventional thermal and laser annealing will be performed and compared. Recrystallization, thermal stress and composition will be analyzed. Film qualities such as refractive index, acoustic and optic loss, electromechanical coupling coefficient, and different electronic parameters will be measured. These measurements are to be correlated with the orientation, size and density of the grain, surface smoothness, color of the film, etc. Surface and interfacial properties of the thin films will be determined by conventional capacitance measurements as well as by SAW techniques. Special emphasis will be placed upon investigating the thin film qualities of AℓN on SOS (Si on sapphire), InSb on $LiNbO_3$, ZnO on GaAs,etc., since these structures may be useful in certain AE and AO devices. Acoustic propagation characteristics as well as acoustoelectric interactions in these thin film layered structures will be studied both in theory and experiments. In the case of AE interactions in layered structures, a set of field-coupled, two-dimensional differential equations will be analyzed with the involvement of surface and interfacial conditions in the thin film structures.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

1) One of our ongoing programs has involved the fabrication of InSb films of superior quality. The films we have been able to fabricate are probably the best in the world outside of the Soviet Union, but they still require improvement. Toward this goal, we are in the process of installing two new stations in our fabrication facility which are to be used exclusively for the deposition of InSb films and the overlay aluminum nitrate (AℓN) films. An ion beam sputtering station for InSb deposition has just been completed and is now being set into operation. The second station for the AℓN films, which are used as the protective overlay for the InSb films, is expected to be in operation soon.

2) In connection with our program on monolithic thin film semiconductor convolvers, we have supplemented our experimental efforts with a theoretical analysis of the transverse component method of operation. The problem is actually three-dimensional and therefore very complicated, but, in the case of InSb on $LiNbO_3$, the InSb layer is sufficiently thin that one can assume that the potential function is independent of the thickness across the InSb film. With that approximation, it should be possible to obtain a solution, and calculations are in progress. On the experimental side of the monolithic semiconductor convolver program, we have during this past year presented a talk[11] at the IEEE Ultrasonics Symposium, and published a journal paper.[12]

3) We have this year, in conjunction with our InSb thin film SAW amplifier program, initiated an investigation of a new type of ultrasonic imager utilizing multistrip track-changer reflectors. This new type of imager has the advantage over the conventional one of not requiring

clock signals. Discussions in the next two sections indicate the structure required and outline the challenges. The project has started only recently, with the successful construction of a single track changer.

4) A monolithic thin film real time correlator has been realized in ZnO on Si form and in ZnO on SOS form. The correlator involves two modes of propagation (the Rayleigh mode and the Sezawa mode) in the ZnO film; their distinctly different velocities permit the generation of the correlation signals in a novel manner. The results obtained so far from this study will be presented[14] at the October 1983 IEEE Ultrasonics Symposium.

5) For some applications, such as acoustoelectric imagers employing ZnO on glass, the ZnO film must be photosensitive. We have found that the photosensitivity of ZnO can be controlled by different annealing conditions, particularly the annealing temperature, where, for example, appropriate annealing produced a change in photoconductivity of five orders of magnitude.

4. STATE OF THE ART AND PROGRESS DETAILS

During the past few years we have made the following major accomplishments.

(1) We have successfully designed and constructed a novel and superior thin film deposition system utilizing a modified sputter gun. This new system permits us to fabricate thin piezoelectric or semiconducting films of exceptionally high quality at both low temperature and low pressure. The new system has many additional advantages: it is cost effective (low cost both to construct and to operate), it has a high deposition rate (comparable to that of the more costly and less advantageous planar magnetron system), it reduces greatly the problem of thermal stresses, and it permits the fabrication of clear, well-oriented, high quality films on many different substrates, in contrast to other systems.

(2) Using our newly-developed sputtering system, we have fabricated excellent thin films of ZnO and AlN on various substrates. Good InSb films have also been made using evaporation techniques. The ZnO films, which came first, are as good as any made elsewhere, to our knowledge, and have received high praise from industry. We have also fabricated them on glass, silicon and quartz substrates, with equal facility and quality. The low pressure associated with our sputtering system permits us to produce AlN films of exceptional quality, colorless and with an extremely smooth surface. The InSb films we have produced are not yet of top quality, but to our knowledge they are the best in the world outside of the Soviet Union, and we have achieved bidirectional SAW amplification with them, which indicates that they are already quite good.

(3) Because of discrepancies which consistently appeared in experimental data involving the acoustoelectric (AE) effect, we theoretically examined the Weinreich relation, which relates the acoustoelectric current to the acoustic power loss and the carrier mobility. A two-dimensional analysis was carried out on a configuration consisting of a semiconductor

film on a SAW substrate, taking into account both the transverse and tangential dc acoustoelectric effects. It was found that the Weinreich relation, which is widely used and assumed to be correct, is not valid most of the time it is used in SAW configurations. Its range of validity is difficult to express simply, but we are currently clarifying it.

(4) Since monolithic thin film semiconductor convolvers are of substantial interest and should become feasible now that very good thin films can be fabricated, we examined the properties of several possible types of such convolvers. The stress in the study was on the semiconductor film on a piezoelectric substrate, which should have a time-bandwidth product comparable to that of the successful but inconvenient separated-medium type. We made calculations on the dispersion, attenuation, and efficiency, as a function of film thickness, conductivity and frequency, for two types of output terminal arrangement. These analyses involved new considerations, not taken into account by available theories. We have recently constructed a new monolithic integrated convolver with transverse biasing, and we have observed strong output enhancement.

(5) In the past year, in conjunction with our InSb thin film SAW amplifier program, we initiated the study of a new type of ultrasonic imager utilizing multistrip track-changer reflectors. This type of scanner has the advantage of operating without the need of clock signals. A single track-changer reflector has been constructed and successfully tested. The goal, of course, is to construct a serpentine scanner utilizing successive track-changer reflectors. The crucial tasks to be achieved are (i) the aperture of the track-changer reflecters has to be small, since the resolution of the imager is inversely proportional to the aperture size, and (ii) the InSb surface acoustic wave amplifier has to be properly incorporated and integrated into the imager, so that the sonic attenuation can be compensated. This project is in progress.

(6) A monolithic thin film real time correlator has been realized utilizing the structure of ZnO on Si (also ZnO on SOS). In the ZnO film both the first mode (Rayleigh) and second mode (Sezawa) were excited. The same mask was used for the generation of both modes. The first mode, with the IDT deposited at the interface of ZnO and Si, is operated at a center frequency of 130 MHz, and the second, with the IDT on the ZnO surface, operated at 220 MHz. Thus, their velocities are distinctly different, with a ratio of 1.7 to 1. The correlation output was taken in the width direction transverse to both the sonic propagation and the substrate normal. Transverse DC field was also applied to enhance the output signal. The design of the correlator is such that both the waves are propagating in the same direction. We let the slow mode be launched first; then, when the fast mode overtakes the slow mode, correlation signals are generated via the acousto-electric interaction between the two modes. A talk on this study will be presented at the 1983 IEE Ultrasonics Symposium.[14]

Some of the details regarding our progress and accomplishments are briefly summarized below.

SECTION II: SOLID STATE ELECTRONICS

A. ZnO Films Grown at Low Temperature by Our Newly Developed "Modified Sputter Gun" System[1]

The deposition of C-axis oriented ZnO films by a sputtering technique was initiated in the mid-1960s.[2] Since then, ZnO films have been considered to be one of the most promising materials for optical waveguiding, acoustoelectric applications, acoustooptic interaction media, and SAW transducers, based on its high electromechanical coupling coefficient and its superior optical qualities.[3-7] However, ZnO films did not gain momentum until the group at Kyoto University, Japan, reported several years ago that high quality ZnO films can be obtained on 7059 glass substrates at a relatively high growth rate ($\cong$ 2 μm/hr) by utilizing planar magnetron sputtering.[8-10] (The thermal expansion coefficients of ZnO and 7059 glass are nearly identical.) It is of interest that Dr. S. Onishi, one of the coauthors of this proposed work unit, was a key member of that group under Professor Shiosaki, and that the work on planar magnetron sputtering was the topic of his doctoral dissertation.

However, the films on other substrates, such as silicon, quartz, etc., are normally not optically transparent due to their mismatch in thermal expansion coefficients. In order to improve the quality of ZnO films on all substrates, we studied these past techniques and recognized that in order to achieve clear films one must essentially avoid the problem of thermal stress caused by electron bombardment. We then found how to avoid such stress by suitably modifying the "research S-gun" sputtering system. Presently, this modified S-gun provides us with films of the highest quality. The reason why will become clear as we review the past sputtering methods.

(1) Using a conventional sputtering unit: the substrate holder is of positive polarity. While the ionized A^+ ions strike the targets (ZnO or Zn) to remove material for deposition, the electrons are bombarding the substrate since it is at positive potential. As a result, the substrate temperature is arbitrarily high and can not be controlled. Films thus obtained are of poor quality.

(2) Utilizing a planar magnetron sputtering system: With this system, the majority of the electrons are trapped in the magnetic field. Nevertheless, the substrate will still be bombarded by some of the electrons, since the substrate holder is at positive potential. As stated earlier, the planar magnetron system has significantly improved the film deposition technique. However, ZnO films which are well oriented, clear and of high quality are usually difficult to obtain on other than the 7059 glass substrate.

With the so-called "research S-gun" sputtering system, it was necessary for us to augment and specially arrange the magnetic field so that the thin-film surface is completely outside of the plasma column and free from electron bombardment. The sputtering system is shown in Fig. 1, where an additional magnet is used to shape the plasma column and deflect the electron beam leakage. When these modifications have been made, the S-gun system is the system that has provided us with the best quality of film on various substrates, for the following reasons:

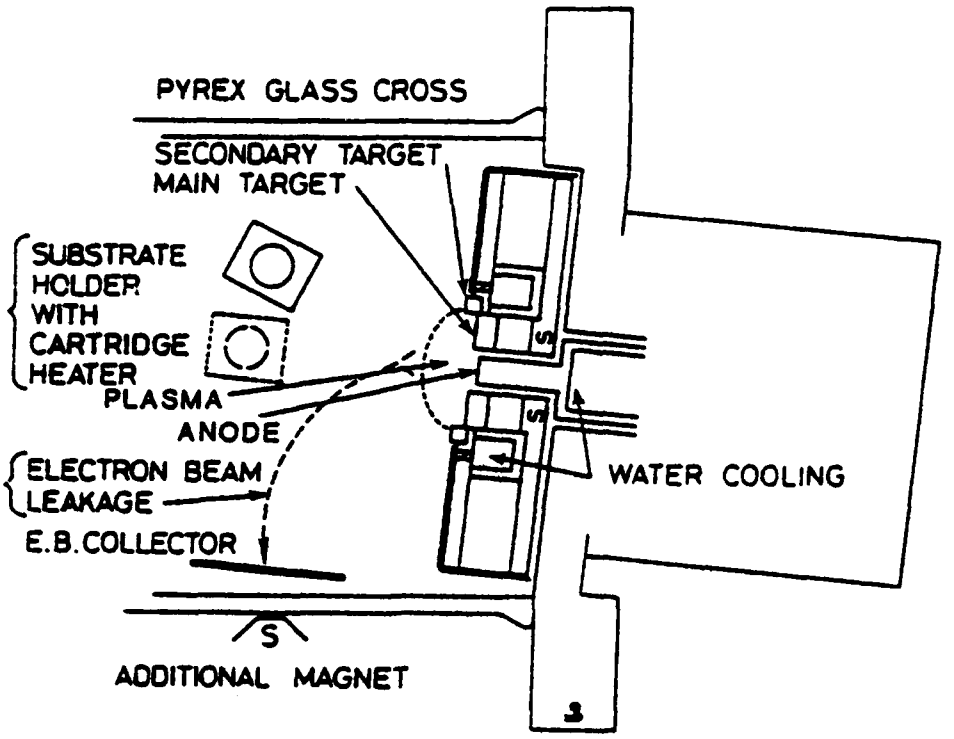


Fig. 1 Geometry of modified sputter gun system.

(a) The substrate holder is at neutral potential, and the substrate surface is free from electron bombardment.

(b) The water cooling system is very effective, since it is right next to the target. The target will therefore be at a lower temperature and the substrate surface will not be subjected to excessive heat radiation from the target.

(c) The magnetic field is several times stronger than that associated with a planar magnetron. The plasma column is therefore more tightly confined.

(d) The deposition rate of an S-gun system is comparable to that of a planar magnetron, whose deposition rate is good.

Therefore, the substrate holder is at a low temperature, around 150°C, rather than about 500°C as with the planar magnetron. This feature is of key significance, since it greatly reduces the problem of thermal stresses. As a result, with this system, we are able to obtain clear,

well-oriented, high quality ZnO films on all substrates available to us, such as glass, silicon and quartz substrates.

In addition to the key features of low pressure and low temperature sputtering, other advantages of this newly-developed sputter gun system include: (a) a much larger area of high quality thin film can be obtained, and (b) the system is cheaper to construct and cheaper to run.

In certain applications, such as ZO on Si acousto-electric devices, the ZO film is required to be of very high resistivity and highly transparent, and in other cases, such as ZO on glass acousto-electric imagers, the ZO film should be photo sensitive. We have found that the ZO photo sensitivity can be controlled by different annealing conditions, especially the annealing temperature. For instance, a 5 orders of magnitude change in photoconductivity (between dark and UV light) has been achieved for samples under an annealing temperature $\sim$ 100°C and an annealing duration of about 5 minutes.

B. AℓN Films Grown at Low Pressure and Low Temperature, Utilizing Our "Modified Sputter Gun" System

The utilization of our modified sputter gun has also been extended for the deposition of aluminum nitride (AℓN) films on various substrates.[11-14] The strong magnetic field associated with this gun system makes it possible to maintain a stable plasma at a total sputtering gas pressure (TSGP) as low as 1 μTorr. This low-pressure sputtering provides films with denser and finer grains and it improves c-orientation, surface smoothness, film color, transparency and deposition rate. Our experience shows that the qualities of AℓN films are extremely sensitive to the level of total sputtering gas pressure, the lower the better. The total pressure in our system varies from 1 μTorr to 8 μTorr. The deposition rate of the sputter gun is three times as high as that of a conventional planar magnetron, since the target cooling system makes it possible to increase rf or dc power. It is 2.4 μm/hr. at an input power of 300 watts. The experimental results in Fig. 2 show that the growth rate decreases with increase in total pressure. And the film is only transparent and colorless at low pressure sputtering. High quality films are deposited at temperatures around 350°C, which is low for AℓN deposition. The substrate surface in this system, as stated in A, is free from charged particle bombardment and excessive heat radiation.

The acoustic properties of the AℓN films have been revealed by evaluating delay line characteristics.

(i) The insertion loss of an AℓN on SOS delay line is about 20 dB at a center frequency of 250 Mhz, which indicates that the film is of excellent quality and that the piezoelectrical coupling coefficient approaches that of a single crystal. The design parameters for the delay line are:

propagation distance, 1 cm
IDT finger pairs, 12
IDT aperture, 0.1 cm
AℓN film thickness, 4 μm
Si film thickness, 1 μm

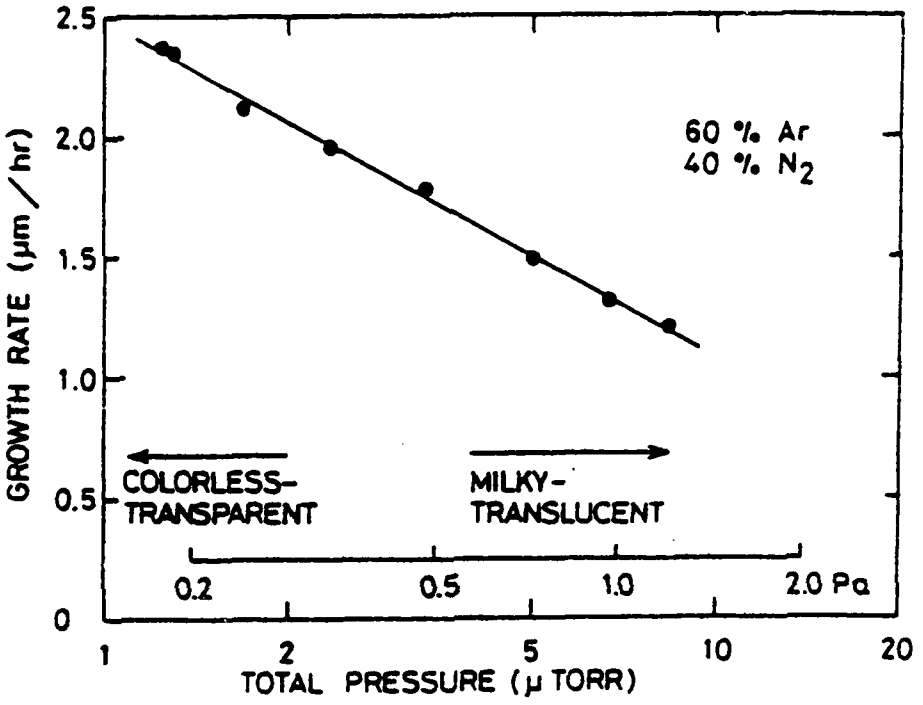


Fig. 2 Deposition condition of AℓN film: total sputter gas pressure vs. growth rate.

(ii) Dispersion characteristics: Velocity dispersion characteristics for the AℓN-Si-sapphire structure have been obtained from delay lines fabricated with different AℓN film thicknesses. As expected, the characteristics of velocity dispersion fall between those of AℓN and those of sapphire.

C. InSb Films

The deposition of InSb thin films on $LiNbO_3$ substrates has been accomplished at different centers since the early 1970s,[15-18] and the process has been used with little variation in Japan, France, Russia and the U.S. However, due to difficulties in obtaining high quality films, acoustoelectric devices using InSb films have not been proven to be practical. For AE applications it is required that the InSb film possesses high drift mobility and high resistance. The best data reported to date are by Kotelyanskii, et al.,[18] in Russia; they can obtain a film of 500Å thickness and a drift mobility of 1600 cm^2/v-sec. This mobility is over three times better than what has been reported in the western world. We have been engaging in the development of InSb films for three years and have achieved bidirectional SAW amplification with those films necessary for reducing the insertion loss of a convolver.[19] The drift mobility we obtained here is $\cong$ 700 cm^2/v-sec at a film thickness $\sim$ 500Å, which is better than what is reported in this country but is still low in comparison with that of Kotelyanskii, et al.[18] Private discussions with Kotelyanskii during his recent visit to us here revealed that their set-up is much more elaborate and costly, and they

have invested 10 years in this effort.

Figure 3 shows some of our experimental results of Hall mobility vs. the conductivity thickness product. The film thickness is about 450Å. Our process, especially on forming the electric contacts, we feel is superior.[19] The highest Hall mobility obtained by us is about 2500 cm/volt-sec, which means that if the traps can be eliminated in this film a drift mobility of that same magnitude can be achieved. The significance of these measurements is that they tell us how good a film can be achieved ultimately, when the Hall and drift mobilities are equal. Thus, these measurements offer us substantial hope.

We have reason to believe that our InSb film quality will be greatly improved, since we are in the process of installing two new stations which are to be used exclusively for the deposition of InSb films and its overlay aluminum nitrate (AℓN) film. The ion beam sputtering station for InSb deposition has been just completed and is now set into operation. The quality of the film after a period of adjustment and experimentation is expected to be superior to previously produced films. The second station for AℓN films used as a protective overlay for the InSb film is expected to be in operation soon.

D. Examining the dc Acoustoelectric Current Produced by SAW

It is generally believed that surface acoustic waves will be useful in determining the surface and interfacial properties of semiconductor wafers and semiconductor films; however, in the process of our correlating data from SAW measurements with those obtained from conventional methods, some discrepancies consistently appear.[20-24] This led us to examine the validity of the commonly-used Weinreich relation for a configuration consisting of a semiconductor on a SAW substrate. The Weinreich formula relates the acoustoelectric current to the acoustic power loss and the carrier mobility. Since both the AE current and the power loss are measurable quantities, the carrier mobility can thus be determined. Comparing the mobility measured here with the Hall mobility measurement, trapping dynamics can also be revealed.

The Weinreich relation was first derived for phonon-drag phenomena in Ge due to the deformation potential.[25] Its validity in bulk piezoelectric semiconductors and in the presence of acoustic gain was experimentally verified by one of us (W.C. Wang) in 1962,[26] and subsequently analyzed by H.N. Spector.[27-28] In examining the Weinreich formula for thin films, however, we have carried out a two-dimensional analysis for a configuration consisting of a semiconductor film on a SAW substrate, taking into account both the transverse and tangential dc AE effects. The analysis shows that the range of validity of the Weinreich relation in such a system is very limited and that in general it cannot be applied here. (Its range of validity cannot be simply stated, but we are currently assessing the validity conditions.) In addition, the analysis reveals that (a) the field associated with the transverse AE current in general tends to accumulate the charge at the semiconductor surface, (b) the polarity of the transverse AE current will, in general, not change sign when the film thickness is larger than the SAW wavelength, and (c) the diffusion effect cannot be ignored in the calculation of tangential AE current.

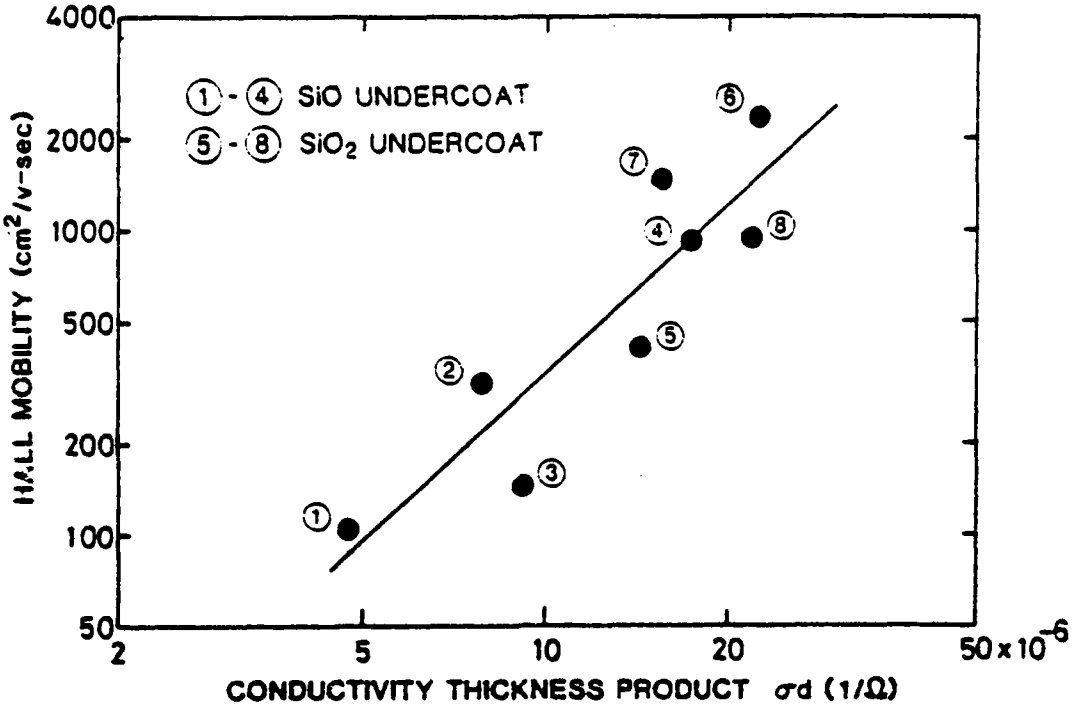


Fig. 3 Hall mobility of InSb film (450Å thick) versus conductivity thickness product.

E. Monolithic Thin Film Semiconductor Convolver

Many forms of convolver have been designed and tested.[29-31] The best performance to date is the one given by the separated-medium Si on $LiNbO_3$ normal-component convolver fabricated at Lincoln Laboratory and Texas Instruments. However, looking ahead and taking into consideration eventual cost effectiveness and smaller size, the monolithic convolver still remains a strong candidate. There are three possible monolithic AE convolver structures: (i) piezoelectric film on semiconductor substrate; (ii) semiconductor film on piezoelectric substrate; (iii) piezoelectric film on semiconductor film. The time bandwidth (TB) product for case (i) is usually rather low due to the dispersion introduced by the piezoelectric film (operating in the Sezawa mode). The TB product for case (ii) should be comparable to that of the separated-medium type. Therefore, we started to analyze its performance. Case (iii) will be investigated in the future. Further remarks about these possible structures are presented in the Proposed Research Program section.

Two types of output terminal arrangement for case (ii) have been investigated. In one type, the output terminal plate is in direct ohmic contact with the top of the semiconductor surface, and in the other an insulating dielectric layer is inserted between the semiconductor surface and the convolver's output plate. Therefore, the former has a short-circuit condition and the latter an open-circuit condition on the semiconductor surface. Since the boundary conditions are different, the

velocity, attenuation constant and convolver efficiency are found to be quite different in the two configurations. Calculations on the dispersion, attenuation and efficiency have been carried out as a function of film thickness, conductivity and frequency.

We should point out that another type of convolver, the elastic convolver, has had much attention lately, and has shown impressive progress. These elastic convolvers are most useful in lower performance applications, where the TB product requirements are not very high. We believe that, even if the elastic convolvers become even better, a semiconducting film will later be placed on top of the elastic convolver configuration (but under the metal cover, of course) to further improve its performance. The ultimate monolithic convolver form will therefore, in our opinion, be a combination of elastic and semiconductor forms. From that standpoint, our studies reported above, and those we propose later, will be relevant independent of the potential success of the elastic convolvers.

We have recently obtained new results on an active monolithic convolver band on an InSb film on $LiNbO_3$. The convolver was operated in two different modes.

(1) Transverse Component Operation:

Recently, we have been engaged in a new experimental development. We have observed enhancement of the output of a thin film InSb/$LiNbO_3$ convolver by transverse biasing. In the design, an integrated and combined thin film InSb on $LiNbO_3$ convolver and amplifier device was constructed. The device was operated successfully at 150 MHz, with an insertion loss around 6 dB instead of the customary 40 dB to 60 dB.

The convolver output, which was taken in the width direction, transverse to both the sonic propagation and the substrate normal, was greatly enhanced by the application of a transverse (width) drift d.c. field on the InSb film. Since the width of the InSb film can be made narrow, the d.c. voltage required is not excessive. The relationship between the convolver output and the transverse d.c. field is approximately linear.

Strictly speaking, the problem is three-dimensional, so that the exact solution is too complicated to obtain. However, in the case of InSb on $LiNbO_3$, the thickness of the InSb layer is only 500 Å, which is about one tenth the Debye length, so that one can assume that the potential function in independent of the thickness across the InSb film. On that basis, a solution can be obtained; calculations are in progress.

(2) Tangential Component Operation:

A detailed theoretical analysis has been carried out for the tangential mode thin film convolver. In the convolver operation both the forward and backward waves are amplified, i.e., the SAW amplifier is bidirectional. The same InSb film strip is used for amplification and convolution in the device geometry. Calculations on the convolver efficiency as a function of drift field, film thickness and conductivity have

been made. Since the geometry of the multilayer structure, InSb on $LiNbO_3$ with a very thin passivation layer of SiO_2 sandwiched in-between, is very similar to that of a separated medium amplifier and convolver, the technique used in the analysis is quite similar to that for a separated medium one. However, it is much more involved here since bidirectional gain has to be considered. Experimentation on the bidirectional tangential mode active convolver was also carried out; the convolver efficiency has been found to be far less than that of the transverse mode convolver.

5. REFERENCES

1. S. Onishi, M. Eschwei and W.C. Wang, Appl. Phys. Lett. 38, 419 (1981).

2. S. Wanuga, T.A. Midford and J.P. Dietz, IEEE Ultrasonics Symposium (1965).

3. F.S. Hickernell, J. Appl. Phys. 44, 1061 (1973).

4. F.S. Hickernell and J.W. Brewer, Appl. Phys. Lett. 21, 38[illegible] (1972).

5. R.S. Wagers, G.S. Kino, P. Galle and D. Winslow, IEE[illegible] Ultrasonics Symposium (1972).

6. R.F. Pierret, R.L. Gunshor and M.E. Cornell, IEEE Ultrasonics Symposium (1979).

7. S. Furukawa, S. Tsuzihara, T. Moriizumi and T. Yasuda, IEEE Ultrasonics Symposium (1979).

8. T. Shiosaki, S. Onishi and A. Kawabata, J. Appl. Phys., 50, 3113 (1979).

9. T. Onishi, Y. Hirokawa, T. Shiosaki and A. Kawabata, Japanese JAP 17, 773 (1978).

10. F.S. Hickernell, IEEE Ultrasonics Symposium (1980).

11. S. Onishi, M. Eschwei, S. Bielaczy and W.C. Wang, to appear in Appl. Phys. Letters (October 15, 1981).

12. A.J. Noreika, M.H. Francombe and S.A. Zeitman, J. Vac. Sci. Technol. 6, 194 (1969).

13. A.J. Shuskus, T.M. Reeder and E.L. Paradis, Appl. Phys. Lett. 24, 155 (1974).

14. T. Tsubouchi, K. Sugai and N. Mikoshiba, Proc. IEEE Ultrasonic Symp. 446 (1980).

15. L.A. Coldren, Appl. 18, 319 (1971).

16. L.A. Coldren and G.S. Kino, Appl. Phys. Lett. 23, 117 (1973).

17. K. Yamanouchi, K. Abe and K. Shibayama, Suppl. J. Japan Soc. Appl. Phys. 43, 203 (1974).

18. L. M. Kotelyanskii, A. I. Krikunov, A. V. Medved, R.A. Mishkinis and V.V. Panteleev, Sov. Phys. Semicond. 12, 751 (1978).

19. M. Eschwei, S. Onishi, F. Cassara and W.C. Wang, to be presented at American Vacuum Society, 18th National Vacuum Symposium (1981).

20. K.S. Chen, Z.S. Wu, H. Schachter and W.C. Wang, to be presented at IEEE Ultrasonics Symposium (1981).

21. J.H. Cafarella, A. Bers and B.E. Burke, IEEE Symposium Proceedings, 181 (1972).

22. A. Bers, J.H. Cafarella and B.E. Burke, Appl. Phys. Lett. 22, 399 (1973).

23. R. Adler, D. Jones, S. Datta and B.J. Hunsinger, IEEE Ultrasonics Symposium (1980).

24. P. Das, R.T. Webster, H. Estrada-Vazquez and W.C. Wang, Surface Science 86, 848 (1979).

25. G. Weinreich, Phys. Rev. 107, 317 (1957).

26. W.C. Wang, Phys. Rev. Lett. 9, 443 (1962).

27. H.N. Spector, J. Appl. Phys. 34, 3628 (1963).

28. K.A. Ingebrigtsen, J. Appl. Phys. 41, 454 (1970).

29. W.C. Wang and P. Das, IEEE Ultrasonics Symposium (1972).

G.S. Kino, W.R. Shreve and H.R. Gautier, IEEE Ultrasonics Symposium (1972).

J.M. Smith, E. Stern, A. Bers and J. Cafarella, IEEE Ultrasonics Symposium (1973).

30. L.P. Solie, Proc. IEEE 64, 760 (1976).

31. P. Defranould and C. Maerfeld, Proc. IEEE 64, 748 (1976).

6. RECENT PUBLICATIONS

1. L. Rosenheck, H. Schachter and W.C. Wang, "Acoustoelectric Real Time Correlator," Japanese Journal of Applied Physics, Vol. Supp. 18-1, p. 215 (1979).

2. P. Das, R. Webster, H. Estrada and W.C. Wang, "Contactless Semiconductor Surface Characterization Using Surface Acoustic Waves," Surface Science, Vol. 86, p. 848 (1979).

3. W.C. Wang, H. Schachter and F. Cassara, "Acousto-Electric Surface Acoustic Wave Demodulators," Proc. IEEE Ultrasonics Symposium, Cat. #79-CH-1482-9SU (1979).

4. S. Onishi, M. Eschwei and W.C. Wang, "Transparent and Highly Oriented ZnO Films Grown at Low Temperature by Sputtering with a Modified Sputter Gun," Applied Phys. Lett., Vol. 38(6), p. 319 (1981).

5. S. Onishi, M. Eschwei, S. Bielaczy, and W.C. Wang, "Colorless-Transparent, c-Oriented Aluminum Nitride Films Grown at Low Temperature by a Modified Sputter Gun," Applied Phys. Lett., 39, 643 (1981).

6. M. Eschwei, S. Onishi and W.C. Wang, "Properties of InSb Film for SAW Amplifier," accepted for presentation in American Vacuum Society 28th National Vacuum Symposium (1981).

7. Z.S. Wu, S. Onishi, K.S. Chen and W.C. Wang, "Monolithic Thin Film Acoustoelectric Convolver," to be presented in 1981 Ultrasonics Symposium.

8. S. Onishi, M. Eschwei, S. Bielaczy and W.C. Wang, "Acoustic, Optical Properties of AlN Films Grown at Low Temperature by Modified Sputter Gun," to be presented in 1981 Ultrasonics Symposium.

9. K.S. Chen, Z.S. Wu, H. Schachter and W.C. Wang, "D-C Acoustoelectric Current Induced in a Semiconductor Film Adjacent to a SAW Substrate," to be presented in 1981 Ultrasonics Symposium.

10. M. Eschwei, S. Onishi, F. Cassara and W.C. Wang, J. Vac. Sci. Technol., 20, 873 (1982).

11. S. Onishi, M. Eschwei, B. El-Asir, S. Bielaczy, H. Schachter and W-C. Wang, "Enhancement of Monolithic Convolver Output by Transverse Biasing," presented at IEEE Ultrasonic Symposium (October 1982).

12. S. Onishi, M. Eschwei, B. El-Asir and W-C. Wang, "Enhancement of a Thin Film Indium Antimonide on $LiNbO_3$ Convolver by Transverse Biasing," Japanese J. of Appl. Phys., 22, 273, (May 1983).

13. F.A. Cassara and W-C. Wang, "Laboratory Experiments in SAW Devices and their Applications," IEEE Trans. on Ed., 26, 52 (May 1983).

14. S. Onishi, B. El-Asir, M. Eschwei and W-C. Wang, "Monolithic Thin Film Acousto-Electric Correlator Utilizing Two Modes of SAW," to be presented at IEEE Ultrasonics Symposium (October 1983).

7. DoD AND OTHER INTERACTIONS

Industry has expressed substantial interest in our thin film fabrication capability and its potential for producing excellent films in a cost-effective manner. Two examples are given:

(a) We have a contract with Fairchild Camera and Instrument for a program which would utilize an InSb thin film amplifier in the construction of a compact and low loss SAW delay line for repeater-jammer applications.

(b) The earliest public recognition of our capability was made by Dr. F. Hickernell of Motorola, Arizona, with whom we were in contact. During his invited talk on ZnO thin films at the 1979 IEEE Ultrasonics Symposium, he held up one of our high quality (transparent and colorless) ZnO films on glass and stated that this is an example of what a really good film should be like.

In addition, we have had visits from a large variety of sources. Some examples are:

(c) Drs. J.M. Kotelyanskii and V.T. Potapov of the Institute of Radioengineering and Electronics, Academy of Sciences of the USSR, Moscow, visited us in Brooklyn (and therefore did not see our facility, which is in Farmingdale), and we exchanged information on InSb films. Dr. Kotelyanskii's laboratory produces the best InSb films in the world. It is of interest that about one month later, at an URSI General Assembly in Washington, DC, Professor Oliner was approached by a "young scientist" delegate, Dr. V.I. Anisimkin, from the same laboratory in the USSR, who referred to that meeting and asked if he could spend up to six months in Professor Wang's laboratory.

(d) Professors P. Kornreich and S. Kowel of Syracuse University visited us to discuss our AℓN and ZnO deposition techniques.

(e) Dr. H. Gautier and Dr. P. Defranould of Thomson-CSF, Cagnes sur Mer, France, visited us to see our fabrication facility.

(f) Among others who have visited our facility are Mr. E. Stern of the MIT Lincoln Laboratory and Professor P. Das of the Rensselaer Polytechhnic Institute.

SECTION III
INFORMATION ELECTRONICS

A. ADAPTIVE FILTERING AND SPECTRAL ESTIMATION

Professor A. Papoulis

Unit IE3-1

1. OBJECTIVE(S)

The general objectives remain essentially the same:

(a) Study of the properties of adaptive frequency domain filters with applications in estimation and detection theory.

(b) Spectral estimation with emphasis on extrapolation techniques.

(c) Applications of the method of maximum entropy under prior constraints including Bayesian methods.

(d) Estimation of line spectra from noisy data.

2. APPROACH

In the study of adaptive filters, we concentrate on the determination of the statistical properties of the Widrow filter designed either directly in the time domain or in terms of running transforms, and we investigate the effect of round-off errors, stability considerations involving periodic sources, and conditions for asymptotic equilibrium. In the solution, we use an adaptation of Floquet theory to recursion equations and perturbation techniques involving small departures from equilibrium.

In the extrapolation problem, we reexamine the bandlimited extrapolation method developed earlier[13,14] assuming that the data are determined from a single sample of the process. We further determine the class spectra that are consistent with the constraints and we select a particular form that meets prior information and leads to simple computations involving ARMA models (for details see Section 9).

We extend the method of maximum entropy to problems where the data involve the values of a single sample of the unknown spectrum and we seek to determine the asymptotic properties of the resulting spectra as the length of the sample increases.

We note below selectively some of the proposed approaches. Details of the early results are given in the cited papers.

The problem of hidden periodicities has been considered by Pisarenko[10] and others in the context of a result due to Caratheodory. We propose a method that utilizes Levinson's algorithm and the properties of the prediction error filter. In the method of maximum entropy, we assume that the autocorrelation $R[m]$ of the process $s[n]$ is known for $|m| \leq N$ and that $S(\omega)$ is an all-pole function:

$$S(\omega) = \frac{1_N}{\left| 1 - a_1^N e^{-j\omega T} \cdots -a_N^N e^{-jN\omega T} \right|^2} \qquad (1)$$

The unknown parameters are determined recursively (Levinson):

$$e_{N-1}\Gamma_N = R[N] - \sum_{k=1}^{N-1} a_k^{N-1} R[N-k] \qquad (2)$$

$$a_k^N = a_k^{N-1} - \Gamma_N a_{N-k}^{N-1} \qquad 1 \le k \le N-1$$

$$a_N^N = \Gamma_N \qquad e_N = (1-\Gamma_N^2)e_{N-1} \qquad (3)$$

In our approach, we assume that the unknown spectrum consists of lines

$$S(\omega) = \lambda_o + 2\pi \sum_{i=1}^{N} \gamma_i \delta(\omega-\omega_i) \qquad (4)$$

and we show that their parameters can be determined by a modification of the above algorithm. We, thus, establish Caratheodory's results using the properties of the error filter

$$E_N(z) = 1-a_1^N z^{-1} - \cdots - a_N^N z^{-N} \qquad (5)$$

This approach is computationally simpler and it leads to a method for estimating the parameters λ_o, γ_i, and ω_i.

In the extrapolation problem, we modify the so-called Papoulis-Gerchberg algorithm,[13,14] first presented in an earlier JSTAC report,[12] to problems involving noise and prior energy constraints.

In the echo-cancelling problem (telephone communications) we observe the sum

$$r[n] = s[n] + y[n]$$

where s[n] is the signal coming from location B and y[n] is an attenuated and distorted version of a signal coming from location A. The purpose of the filter is to remove y[n]. It has been observed that, if the signal s[n] contains periodicities the adaptation might diverge. We are proposing the investigation of the statistical properties of the filter parameters and the conditions for convergence of the adaptation.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contribution can be understood more clearly.

We are dealing with a general class of problems involving running Fourier transforms, adaptive filters, and spectral estimation. These problems have applications in signal processing, geophysics, meteorology, system theory, optics and many other areas. In our investigation, we concentrate primarily on methods of general interest.

Our results involve primarily the deterministic version of these problems, where it is assumed that the data are known exactly either as direct noiseless measurements or as statistical averages obtained as time averages from sufficiently long samples. These results led to a number of publications and papers cited in the following sections.

Our objective in next year's effort is to concentrate the effort on the stochastic version of these problems where the data are noisy or the available samples are limited. These problems are in general difficult, involving often nonlinear time-varying equations with random coefficients. The available results are mostly in the form of asymptotic theorems, and the various methods are often justified empirically. In the proposed research we plan to consider various special cases where the random components of the data are small, and to apply perturbation techniques for their solution.

4. STATE OF THE ART AND PROGRESS DETAILS

Early results are in the form of conference papers, papers accepted for publication and papers in preparation. We note the following.

On bandlimited extrapolation: A modification of the algorithm introduced in ref. 12 has led to the solution of a problem involving the estimation of a bandlimited function subject to prior energy constraints in terms of a finite segment contaminated by noise. The results are reported in ref. 3.

On the adaptive echo-cancelling filter: We considered the familiar algorithm

$$a_k[n] = a_k[n-1] + \mu\, \bar{\varepsilon}\,[n]\, \bar{x}\,[n-k] \qquad (6)$$

where $\bar{x}[n]$ and $\bar{\varepsilon}[n]$ are truncated versions of the data $x[n]$ and the error $\varepsilon[n]$. We assume that the truncation levels for each signal are different, and that the running DFT of the transmitted signal is periodic. We have established necessary and sufficient conditions for stability, but only for this simplified model. A paper is in preparation.

On spectral estimation: A method is being developed for estimating the parameters λ_o, γ_i, and ω_i of the spectrum $S(\omega)$ of a process consisting of periodicities hidden in noise. The method makes use of

the following property of $E_N(z)$ [see (4)]: If in Levinson's algorithm $|\Gamma_m| = 1$ and $|\Gamma_N| < 1$ for all $N < M$ then all roots of $E_M(z)$ are on the unit circle and all roots of $E_N(z)$ for $N < M$ are inside the circle. We shall use this to estimate the optimum N and the unknown parameters in terms of a sample of s[n]. Preliminary results are reported in paper 1.

On FM detection: Our objective is the determination of the spectra of FM signals using the running DFT. As a preparation, we have analyzed such signals, including the determination of the spectra of FSK and other cyclostationary processes.

The cited publications (references 1 to 9) give an indication of the variety of applications of this investigation. We note in particular the following.

We established a theorem relating the entropy rate of the output of a linear system to the entropy rate of the input, and we applied the result to a class of problems involving parameter estimation (reference 1).

We used a modified form of Levinson's algorithm to prove Wold's decomposition of a random process into a predictable and a regular component (reference 2).

We determined the spectral properties of cyclostationary processes, PAM, and FSK signals (reference 3).

Using a modified form of Kolmogoroff's equation in the context of Markoff processes, we determined the spectrum of FM signals whose instantaneous frequencies are Markoff (reference 4).

We examined the problem of detecting spectral lines in the presence of noise using a modified form of Levinson's algorithm, and applied the results to the detection of point and line sources in a noisy image (reference 5).

A modified form of the bandlimited extrapolation algorithm introduced earlier (reference 12) led to the solution of a problem involving the estimation of bandlimited functions subject to prior energy constraints (reference 6).

We applied the method of maximum entropy to a class of deterministic problems involving the determination of unknown parameters from incomplete data (reference 7).

We applied extrapolation techniques of spectral estimation to the three-dimensional x-ray problem of determining ab initio the structure of a molecule (reference 8).

We applied running transforms and adaptive filters to the echo-cancelling problem in telephone communications (reference 9).

We are continuing research in the areas reported above with a concentration on the deterministic and stochastic aspects of spectral estimation. The investigation covers discrete-time and continuous-time

processes in one and two dimensions and focuses on the characterization of classes of spectra that are consistent with specified constraints. We give next, briefly, an idea of the approach, limiting the discussion to the one-dimensional discrete-time case.

We wish as usual to determine the class of spectra

$$S(z) = \sum_{m=-\infty}^{\infty} R[m]z^{-m} \qquad S(e^{j\omega T}) \geq 0 \tag{7}$$

such that R[m] is specified for $|m| \leq N$. This problem has been investigated in the context of the theory of moments[15] and a solution is presented based on concepts from network theory[16,17]. We propose to reexamine it basing the analysis on various extensions of the lattice filter interpreted as a four-terminal device.

In Fig. 1, we show a single section of a lattice characterized in terms of the reflection coefficient Γ_i where

$$|\Gamma_i| \leq 1 \tag{8}$$

Denoting by $x_{i-1}[n]$, $y_{i-1}[n]$ the two inputs, and by $x_i[n]$, $y_i[n]$ the resulting outputs, we have

$$x_i[n] = x_{i-1}[n] - \Gamma_i y_{i-1}[n-1]$$

$$y_i[n] = -\Gamma_i x_{i-1}[n] + y_{i-1}[n-1]$$

Thus, the transfer function of a single lattice is the matrix

$$\begin{bmatrix} 1 & -\Gamma_i z^{-1} \\ -\Gamma_i & z^{-1} \end{bmatrix} \tag{9}$$

Connecting N such sections in cascade, we obtain the system of Fig. 2. The transfer function of this system equals

$$\begin{bmatrix} A_N(z) & B_n(z) \\ C_N(z) & D_N(z) \end{bmatrix} = \prod_{i=1}^{N} \begin{bmatrix} 1 & -\Gamma_i z^{-1} \\ -\Gamma_i & z^{-1} \end{bmatrix} \tag{10}$$

where, as we can see by induction,

$$C_N(z) = z^{-N}B_N(z^{-1}) \qquad D_N(z) = z^{-N}A_N(z^{-1}) \tag{11}$$

and

$$A_N(z)D_N(z) - B_N(z)C_N(z) = P_N z^{-N} \tag{12}$$

$$P_N = \prod_{i=1}^{N} (1 - \Gamma_i^2)$$

Thus,

$$\begin{aligned} X_N(z) &= X_o(z)\, A_N(z) + Y_o(z)B_N(z) \\ Y_N(z) &= X_o(z)\, C_N(z) + Y_o(z)D_N(z) \end{aligned} \tag{13}$$

The familiar lattice filter obtained from Levinson's algorithm is a special case obtained by connecting the two inputs of the four-terminal lattice as in Fig. 3:

$$x[n] = x_o[n] + y_o[n] \tag{14}$$

As it is well known, the transfer functions from the common input to the two outputs equal

$$E_N(z) \quad \text{and} \quad z^{-N}E_N(z^{-1})$$

respectively, where $E_N(z)$ is the prediction error filter (5). Hence [see (14)]

$$E_N(z) = A_N(z) + B_N(z) \tag{15}$$

The constants Γ_i are determined as in (2) and (3) and the resulting spectrum

$$S(z) = \frac{P_N}{E_N(z)E_N(z^{-1})} \tag{16}$$

is such that the values $R[n]$ of its inverse for $|n| \leq N$ agree with the given data.

Suppose now that we extend the lattice beyond the available N (Fig. 4). Denoting by $A_M(z)$ and $B_M(z)$ the upper elements of the continuation, we conclude that the resulting error filter equals

$$E_{N+M}(z) = E_N(z)A_M(z) + z^{-N}B_M(z)E_N(z^{-1}) \tag{17}$$

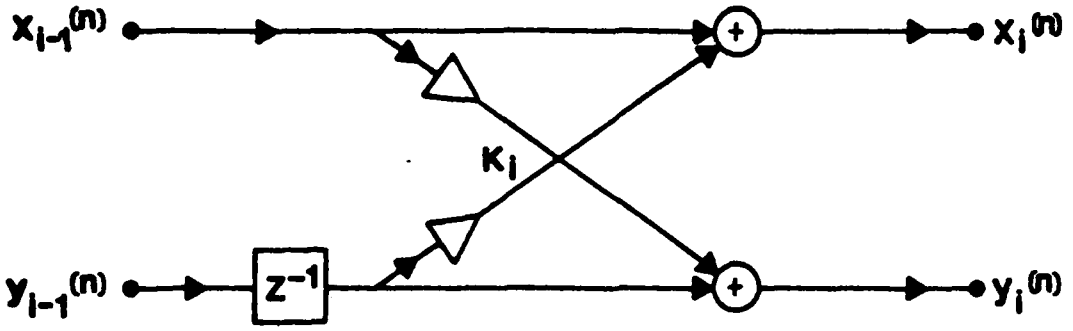


Fig. 1 Section of a lattice filter.

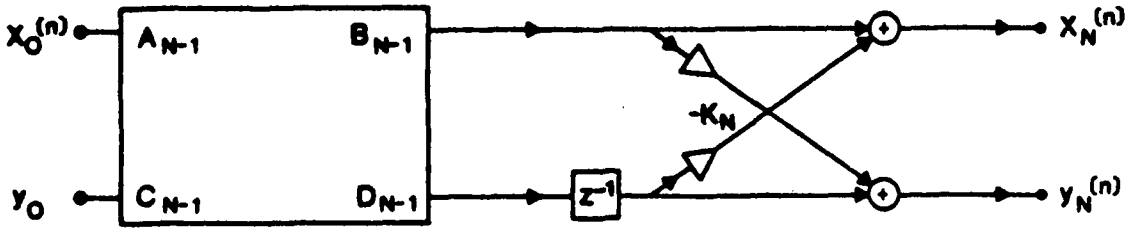


Fig. 2 Lattice as a four-terminal system.

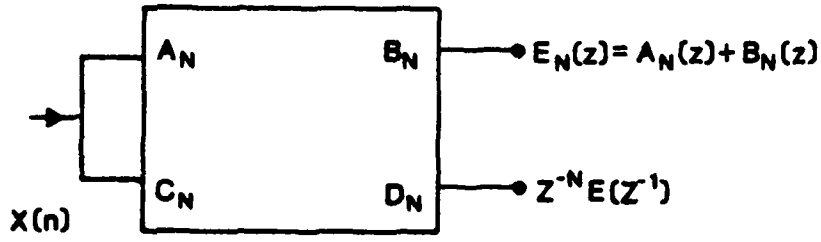


Fig. 3 Lattice as a three-terminal system.

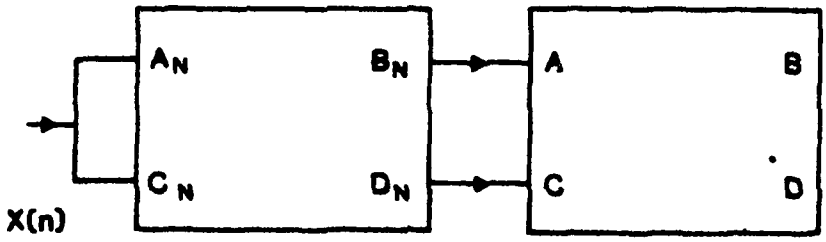


Fig. 4 Cascade of lattice filters.

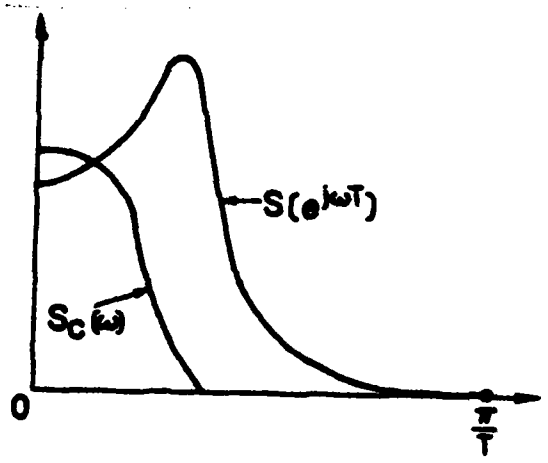


Fig. 5 Effect of oversampling in ME spectral estimation.

In the preceding, $E_N(z)$ is specified in terms of the given data, and the polynomials $A_M(z)$ and $B_M(z)$ are specified as in (10), where Γ_i are such that $|\Gamma_i| \leq 1$ but otherwise arbitrary.

From the preceding it follows that if

$$S(z) = \frac{P_N P_M}{E_{N+M}(z) E_{N+M}(z^{-1})}$$

then its inverse $R[n]$ agrees with the data for $|n| \leq N$, and for $|n| > N$ it gives all possible extrapolations as $N \to \infty$. Thus, the most general spectrum compatible with the data is the function

$$S(z) = \frac{P_N P}{E(z) E(z^{-1})} \tag{18}$$

where

$$E(z) = E_N(z) A(z) + Z^{-N} B(z) E_N(z^{-1})$$

The function $E_N(z)$ is determined in terms of the data using Levinson's algorithm. The functions $A(z)$ and $B(z)$ are specified in terms of the arbitrary coefficients Γ_i. They can also be specified in terms of the following properties [see (11), (12) and (15)]

$$A(z)A(z^{-1}) - B(z)B(z^{-1}) = P \tag{19}$$

$$A(z) + B(z) = E_\infty(z)$$

In the above, $E_\infty(z)$ is an analytic function in the unit circle and P is a constant.

We note finally that $S(z)$ can be written in the form

$$S(z) = \frac{1-\rho(z)\rho(-z)}{Q(z)} P_N \tag{20}$$

where

$$Q(z) = E_N(z) + Z^{-N} \rho(z) E_N(z^{-1})$$

and

$$\rho(z) = \frac{B(z)}{B(z^{-1})}$$

This follows readily from (18) and (19). The above form has a network interpretation[16].

One of the tasks in this investigation involves the selection of $\rho(z)$ so as to satisfy prior requirements. We give below an illustration.

The discrete-time problem of spectral estimation is often obtained by sampling a continuous-time process with bandlimited spectrum $S_c(\omega)$. In this case,

$$S(e^{j\omega T}) = S_c(\omega) \qquad |\omega| < \frac{\pi}{T} = \sigma$$

If the process is oversampled, then $S_c(\omega)$ equals zero in part of the principal interval $(-\sigma,\sigma)$ (Fig. 5). In this case, rational spectra obtained, say, with the maximum entropy method, are poor estimates. In such cases, a preselected extrapolating term $\rho(z)$ improves the estimate.

5. PUBLICATIONS

1. A. Papoulis, "On Entropy Rate" (Invited paper) Third Workshop on Maximum Entropy, University of Wyoming, August 1-4, 1983.

2. A. Papoulis, "From Levinson's Algorithm to Wold's Decomposition" (Invited paper) SIAM 1983 National Meeting, Denver, Colorado, June 6-8, 1983.

3. A. Papoulis, "Random Modulation," IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-31, pp. 96-105, Feb., 1983.

4. A. Papoulis, "Spectral of FM Signals," Proceedings of the 9th Prague Conference on Information Theory, Feb., 1983.

5. A. Papoulis, "Detection of Line Spectra and Point Sources," (Invited paper) Conference on Signal Recovery with Incomplete Information, sponsored by Optical Society of America and supported by U.S. Air Force Office of Scientific Research, Lake Tahoe, January 1983.

6. W. Xu and Christos Chamzas, "On the Extrapolation of Band Limited Functions with Constraints," IEEE Trans. Acoustics and Signal Processing, (to appear).

7. A. Papoulis, "Modeling and Analysis," (Invited paper) Conference on Modeling and Analysis sponsored by IEEE and Italian National Research Council and supported by NATO, Ercie, Italy, October 1982.

8. A. Papoulis, "Spectral Analysis with Applications In X-rays," (Invited paper) sponsored by European Center For Atomic and Molecular Research (CECAM), University of Paris, France, October 1982.

9. A. Papoulis, "Adaptive Frequency Domain Filters and Echo Cancelling," (Invited paper) sponsored by Spanish government, University of Portugal, Porto, Portugal, October, 1982.

6. REFERENCES

10. V.F. Pisarenko, "The Retrieval of Harmonics," Geoph. J. Res., 1973.

11. U. Grenader and G. Szegö, "Toeplitz Forms and Their Applications, Berkeley University Press, 1958.

12. A. Papoulis, "A New Method of Image Restoration," JSTAC Progress Report No. 39, Polytechnic Institute of New York, 1973-1974.

13. R. W. Gerchberg, "Super-Resolution Through Error Energy Reduction," Opt. Acta, vol. 21, No. 9, 1974.

14. A. Papoulis, "A New Algorithm in Spectral Analysis and Band-limited Extrapolation," IEEE Trans. Circuits Syst., vol. CAS-22, No. 9, pp. 735-742, 1975.

15. N.I. Akriezer, "The Classical Moment Problem," Oliver and Boyd Ltd., 1965.

16. D.C. Youla, "The FEE: A New Tunable High-Resolution Spectral Estimation," Technical Note No. 3, RADC, Cont. No. F30602-78-C-0048, 1978.

17. D.C. Youla and N.N. Kazanjian, "Bauer-Type Factorization..," IEEE Trans. on Circuits and Systems, vol. CAS-25, Feb., 1978.

7. DoD AND OTHER INTERACTIONS

We should like to single out three contributions which resulted from JSEP-supported investigations that have recently attracted general interest, and for two of which personalized names were attached.

(a) The Papoulis-Gerchberg method of spectral estimation

This is an iterative algorithm that yields the power spectrum of a signal x(t) in terms of only a segment of x(t). The method is based on the assumption that it reconstructs the unknown portion of the signal by utilizing the available information in the time and frequency axis. The method is particularly effective if the unknown spectrum consists of lines (hidden periodicities).

(b) The Papoulis window

This is an optimum window developed for high resolution spectral estimates based on the earlier Blackman-Tukey techniques.

(c) The generalized sampling expansion

We developed a method for reconstructing a bandlimited signal $x(t)$ in terms of the samples of N linear functionals $y_k(t)$ of $x(t)$ sampled at 1/N the Nyquist rate. Recently, this result was applied by several authors to a variety of problems in signal processing.

These results are relevant to DoD problems and have a variety of applications, including enhancement of diffraction-limited images.

B. ROBUST METHODS IN PATTERN RECOGNITION, IMAGE PROCESSING, AND CLASSIFICATION AND ESTIMATION PROBLEMS

Professor L. Kurz

Unit IE3-2

1. OBJECTIVE(S)

In a paper which reviews twenty years of accomplishments in pattern recognition and image processing, Fu and Rosenfeld[1] state that in the decision-theoretic approach to image processing problems "we are still looking for effective and efficient feature extraction and selection techniques, particularly in nonparametric and small sample situations" [1, p. 365]. The latter class of problems is one of the major thrusts of this proposal. Though Rosenfeld and Fu refer to nonparametric methods, in view of recent developments it would be appropriate to replace the word "nonparametric" by "robust." In the engineering literature the nonparametric property was understood to mean constant false alarm rate (CFAR), usually disregarding the efficiency of the procedure. On the other hand, robust procedures stress efficiency and insensitivity to variations of the underlying noise distributions; they may or may not possess the CFAR property but in all cases a guaranteed upper bound on the false alarm rate is achievable. It should be noted that true robustness over a broad range of variations in environmental conditions is not possible without some form of adaptability which leads naturally to the need for robust estimation procedures which can continuously estimate the parameters required to maintain robustness of operation. Therefore, further work in the area of robust estimation is of prime importance. In particular, the interaction of robust classification or detection procedures with robust estimation techniques in adaptive processes and systems will be studied. The investigation will encompass three philosophically different classes of techniques: methods based on experimental design techniques, methods based on statistical partition and related tests, image reconstruction and related algorithms. As indicated by Fu and Rosenfeld,[1] these techniques have applications in communications (data compression, speech recognition), medicine (diagnosis, abnormality detection), and automation (robot vision). These techniques also have implications in such areas as underwater and earth sounding, radar maps, search radars, remote sensing, identification of human faces and fingerprints, reliability, machine part recognition, etc.

2. APPROACH

Unlike the approaches used by other researchers to some problems in pattern recognition, image processing and classification -- which concentrated in the past on spectral estimation, Bayesian decision-theoretic, linguistic and ad hoc procedures -- the classes of approaches proposed here are essentially statistical with robust properties. Even if the knowledge of underlying statistics of noise is limited, or changed in the processing or acquisition of data, the methodology outlined in this proposal yields satisfactory performance. Thus, the stress in the approach is on the robustness of the statistical methods -- good performance under varying or poorly defined noise conditions. In particular,

the robustizing approaches based on the m-interval polynomial approximation (MIPA) concept also has implications in such areas as robust recursive estimators, robust spectral estimation and system identification and robust p.d.f. and c.d.f. estimation. Improvements in the MIPA approach lead to less partitions and improved approximation per interval. The latter philosophy leads to improvement in performance while decreasing processing time.

3. SUMMARY OF RECENT PROGRESS

This section presents a brief summary of recent progress; more detailed descriptions are contained in the next section in conjunction with the state of the art so that the nature of the contributions can be understood more clearly.

The problem of object detection based on the use of F- and t-statistics has been refined, and the results were submitted to a journal for pre-publication review. Trajectory detection procedures, based on symmetrical balanced incomplete block (SBIB) design masks to generate designs which produce the sharpest results for a preselected level of testing (confidence), were submitted and accepted for publication by the Journal of Computer Vision, Image Processing and Computer Graphics.

The investigation into the area of robust array detection, initiated in reference 33, was extended to more general models. The M-interval approximation (MIPA) approach to robust detection, estimation, classification, pattern recognition and feature extraction was refined and extended to sequential operations of the processors. The performance of the MIPA methodology as compared to the min-max approach suggested by Huber was given for fixed and sequential modes of operation.

Extension of the algorithmic approach to the image reconstruction problem, initiated in reference 18, to two dimensions involving Walsh and other orthogonal functions in the two-dimensional reconstruction process, were studied. The pertinent algorithms were robustized using three principles: fixed structure robustizer, Mann-Whitney preprocessor and MIPA robustizer.

4. STATE OF THE ART AND PROGRESS DETAILS

This work is in part a natural continuation of some of the work supported previously by the JSEP (and builds upon former accomplishments resulting from that support), and in part consists of completely new material. In particular, material involving masking and related methods based on experimental design techniques, and involving some classes of recursive stabilizer algorithms for image reconstruction, are an outgrowth of the previous contract. Material relating to nonlinear partition classifiers follows from former material. On the other hand, other material is completely new, and is only in general terms related to the previous approaches to the problems. In this category we should mention the concept of the m-interval polynomial approximation (MIPA) approach to classification and estimation problems, the statistical approach to object detection by efficient utilization of F- and t-statistics simultaneously, and the use of F-statistics in gradient-type algorithms such as in problems involving the reconstruction of poorly defined image edges and cracks.

The material is grouped into four clusters of problems: 1. Methods based on experimental design techniques.[2-4] Initial efforts in this area were reported in references 5-8. These problems involve the design of appropriate statistical masks which base their sensitivity on the appropriate F-statistic. The problems considered are object, trajectory and edge detection, and image enhancement. 2. Methods based on partition and related statistical tests. A summary of procedures based on partition tests up to the year 1977 is given in reference 9. Extension of these techniques to other forms of partition tests,[10-15] with particular stress on two-dimensional,[11] two-sample (learning with a teacher or supervised learning) and nonlinear statistics, will be considered. A new approach to generation of robust partition classifiers or detectors based on the concept of m-interval polynomial approximation (MIPA) method, introduced by the senior investigator in connection with some classes of estimation[16] and identification[17] problems, represents a particularly interesting and fruitful approach to classification problems in poorly defined or variable noise environment. 3. Image reconstruction algorithms. The proposed algorithms belong to two classes -- recursive stabilizer algorithms for image reconstruction[18] and maximum gradient algorithms based on F-statistics (finding best edge approximations to cracks). 4. Robust recursive estimators. Though the need of robust recursive estimators has been demonstrated in the past,[18-22] a new class of such estimators is introduced here because these do not require inconvenient rank preprocessing and base their robustness and rapid convergence on the MIPA concept. Additional stress is placed on sequential procedures which are particularly useful in poorly defined and changing noise environments.

A. Methods Based on Experimental Design Techniques

(1) Object Detection and Experimental Designs

When ANOVA is applied to image data, a set of F-statistics is generated.[6-8] These statistics are related to confidence regions and can be used to determine if there exist significant row, column effects or interaction between them. Additional information can be incorporated to determine, with a certain degree of confidence, if there exists significant contrast between a pair of rows or columns. Thus, ANOVA transforms an image into a set of F-statistics. The ANOVA statistics are in general insufficient to describe fully the pattern of an image because it confirms only three possible edges: vertical, horizontal and diagonal. Nothing can be concluded if the pattern is described by edge elements other than the three mentioned above. This has been one of the major handicaps in extending ANOVA techniques to scene analysis. However, there exists a pattern which can be described by its ANOVA statistics. Consider the pattern (array) of Figure 1. It is obvious that this pattern has two parts and each part is homogeneous. This leads to the definition of a standard pattern. Let S be an m×m array, B be its subarray which includes columns 1 to k, $1 \leq k \leq m$, and T be the remaining array. Also, let α denote the confidence coefficient. The column and row F-statistics of S, B and T are denoted by FS_c, FS_r, FB_c, FB_r, FT_c and FT_r, respectively. The confidence interval of a contrast between columns k and k + 1 is[23]

$$(c_k, c_{k+1}) = \bar{X}_{\cdot k} - \bar{X}_{\cdot (k+1)} \pm t_{v,1-\alpha/2}(2MS_e/m)^{\frac{1}{2}}.$$

111111000
111111000
111111000
111111000
111111000

Fig. 1

Definition A.1: S is a standard pattern if the F-statistics of ANOVA and t-values of contrasts satisfy: (a) S is heterogeneous, i.e., $FS_c \geq V_\alpha$ and $FS_r < F_\alpha$; (b) B is homogeneous, i.e., $FB_c < F_\alpha$ and $FB_r < F_\alpha$; (c) T is homogeneous, i.e., $FT_c < F_\alpha$ and $FT_r < F_\alpha$; (d) contrast exists between columns k and k + 1, i.e., (c_k, c_{k+1}) does not cover zero.

For example, the statistics of the array of Figure 2 are FS_c = 105.366, FSr = .935, FB_c = .727, FB_r = .848, FT_c = 1.369, FT_r = .867. The threshold values corresponding to 95% confidence level of S(9×16), B(9×5) and T(9×11) are TS_c = 1.751, TS_r = 2.016, TB_c = 2.687, TB_r = 2.265, TT_c = 1.951 and TT_r = 2.057. The contrast confidence interval is (c_5, c_6) = - 1.934 and -2.288. By comparing the calculated statistics and the threshold values, it is obvious that all four conditions are satisfied. This is a standard pattern of 9 rows and 16 columns with the border occurring between columns 5 and 6 with a 95% confidence level. A standard pattern is a function of four variables: the row size (m) and the column size (n) of the array, the border column (k) in the B subarray and the confidence coefficient (α). A given set of values (m,n,k,α) describes a group of standard patterns for which the confidence levels are equal or higher than α. This is true because α defines a limit and F-values are monoincreasing in nature. All patterns in such a set produce the same visual effects, the only difference being the relative intensity of two subarrays.

In order to apply ANOVA to an array of arbitrary patterns, one would transform the original array into a standard pattern. However, the transformation rules must be derived in a random fashion. Randomization process destroys any a priori knowledge of a given model and, therefore, provides observations free of bias. Let W be such a set of transformations which is one-to-one and onto; namely, $W:S(m\times n) \rightarrow R(p,q)$, where $mn = pq$. One can apply W to any array of size m×n and then analyze the transformed array by ANOVA to see if the original array is transformable to a standard pattern. We can now define the visual equivalence of two arrays.

Definition A.2: Let arrays A and B of the same size m×n and W be a set of randomly generated transformation rules such that $W:A(m\times n) \rightarrow A'(p\times q)$; $W:B(m\times n) \rightarrow B'(p\times q)$. A and B are visually equivalent, with a confidence coefficient α, if both A′ and B′ are standard patterns.

Note that visual equivalence is based on the statistics of the transformed array, not the original one. Since all standard patterns

characterized by the same set of parameters (p,q,k,α) give the same visual effects and the transformation process is one-to-one, one expects the original arrays to exhibit visual resemblance.

4444466666676666
4444467667666666
4445466676676666
5444466766666666
4444466766666667
4444466766766676
4444466766666666
4444466666666666
4444466666666666

Fig. 2

Consider the problem of object detection in the framework of supervised learning. The task is to design a classifier with the knowledge obtained from a given set of samples called teachers or templates. Each template is composed of a model object, the target and its surroundings, the background. The background portion is needed to define the boundary and to increase the relative intensity of the target. Figure 3 shows a proposed recursive learning scheme utilizing ANOVA, randomized transformation rules and the standard pattern. The criteria of terminating the recursive process are the F and t statistics associated with a predetermined confidence level. As shown in the figure, the learning process starts by "coloring" the three components in a template array: target, background pixels and those not to be included in the transformation. The next step is to propose the size of a standard pattern by finding the common denominators of the total number of target pixels and background pixels. Let d_i be such a common denominator or the product of a group of nonrepeating common denominators. Also, let n_t and n_b be the total number of target and background pixels, respectively. Then $p = d_i$, $q = (n_t + n_b)/d_i$ and $k = n_b/d_i$. Even though any d_i can be justified as the final choice of p, it is advised to select d_i to minimize the difference between p and q. In step 3, every target pixel is assigned to a position in the target subarray of the standard pattern. The rule to follow is that the assignments should be one-to-one and onto. To ensure randomness, a random number generator is employed to relate the "from" and "to" pixels. All successful mappings are recorded as transformation rules. The assignments of background pixels to the background subarray are done similarly. In step 4, the transformation rules are applied to the original array and the resultant array is analyzed by ANOVA. If the F and t statistics do not qualify to be a standard pattern, it is clear that the chosen background does not provide enough visual contrast to make the target stand out. A different template or a different set of background pixels should be tried. The process is then repeated from the first step. If the first template is found to be transformable to a standard pattern of confidence level α, the learning process is completed. Figure 4 shows a template, the transformation rules, the standard pattern and all the statistics associated with them for the confidence level 95%.

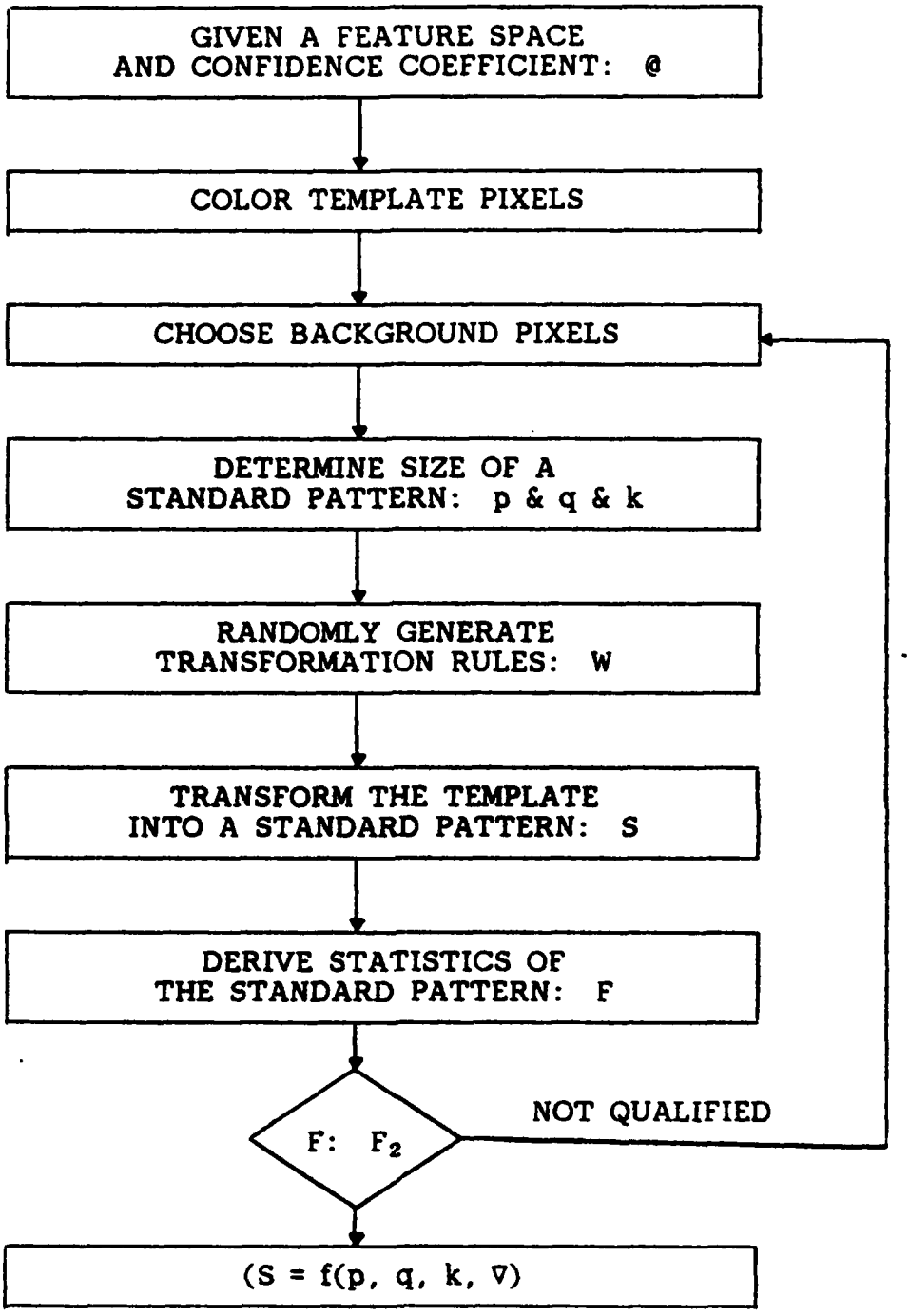


Fig. 3. Supervised learning scheme.

** THE TEACHER FILE ::

1111111111
1111112111
1133354311
1113343311
1133333311
1133353311
1123331111
1113533111
1111111111
1111111211

** THE TRANSFORMED ARRAY **

1111111353
1111111333
1111111333
1112111333
1111121334
1111111435
1112111333
1111111333
1111111353
1111111333

** THE TRANSFORMATION RULES **

(1, 1)-(5, 7) (1, 2)-(5, 1) (1, 3)-(5, 4) (1, 4)-(1, 1) (1, 5)-(1, 3)
(1, 6)-(3, 7) (1, 7)-(9, 7) (1, 6)-(2, 7) (1, 9)-(10, 5) (1,10)-(7, 4)
(2, 1)-(4, 7) (2, 2)-(10, 2) (2, 9)-(1, 5) (2, 4)-(5, 3) (2, 5)-(2, 1)
(2, 6)-(10, 3) (2, 7)-(7, 4) (2, 8)-(9, 4) (2, 9)-(2, 8) (2,10)-(8, 7)
(3, 1)-(1, 7) (3, 2)-(2, 3) (3, 6)-(8,10) (3, 4)-(2, 9) (3, 5)-(4, 5)
(3, 6)-(1, 9) (3, 7)-(5,10) (3, 8)-(8, 8) (3, 9)-(2, 5) (3,10)-(7, 8)
(4, 1)-(4, 5) (4, 2)-(8, 5) (4, 2)-(8, 8) (4, 4)-(10, 9) (4, 5)-(5, 6)
(4, 6)-(6, 8) (4, 7)-(10, 9) (4, 8)-(10,10) (4, 9)-(3, 9) (4,10)-(10, 4)
(5, 1)-(10, 7) (5, 2)-(8, 1) (5, 3)-(9,10) (5, 4)-(1, 8) (9, 5)-(3, 8)
(5, 6)-(7, 9) (5, 7)-(2,10) (5, 8)-(1,10) (5, 9)-(8, 2) (5,10)-(8, 2)
(6, 1)-(3, 1) (6, 2)-(3, 2) (8, 9)-(7,10) (6, 4)-(6, 9) (6, 5)-(9, 1)
(6, 6)-(9, 9) (6, 7)-(2, 8) (8, 8)-(4, 8) (6, 9)-(7, 3) (6,10)-(8, 4)
(7, 1)-(7, 2) (7, 2)-(9, 3) (7, 3)-(5, 8) (7, 4)-(4, 9) (7, 5)-(5, 8)
(7, 6)-(3,10) (7, 7)-(3, 5) (7, 8)-(8, 3) (7, 9)-(8, 4) (7,10)-(6, 5)
(8, 1)-(8, 2) (8, 2)-(1, 4) (8, 3)-(6, 5) (8, 4)-(8, 9) (8, 5)-(6, 3)
(8, 6)-(7, 8) (8, 7)-(3, 9) (8, 8)-(5, 5) (8, 9)-(8, 4) (8,10)-(9, 3)
(9, 1)-(1, 8) (9, 2)-(9, 2) (9, 3)-(9, 8) (9, 4)-(8, 4) (9, 5)-(4, 8)
(9, 6)-(9, 1) (9, 7)-(5, 2) (9, 8)-(2, 1) (9, 9)-(2, 2) (9,10)-(3, 5)
(10, 1)-(8, 7) (10, 2)-(10, 1) (10, 3)-(10, 4) (10, 4)-(4, 6) (10, 5)-(7, 5)
(10, 6)-(7, 6) (10, 7)-(4, 2) (10, 8)-(4, 4) (10, 9)-(1, 3) (10,10)-(4, 3)

: THE CONTRAST BETWEEN COLUMNS :

COLUMN 1	COLUMN 2	CONTRAST
1	2	0
2	3	0
3	4	0
4	5	0
5	6	0
6	7	0
7	8	1
8	9	1
9	10	9

	S_c	S_r	B_c	B_r	T_c	T_r
OBSERVED:	73.553	0.776	1.519	0.818	0.558	1.027
THRESHOLD:	1.999	1.999	2.040	2.040	2.460	2.460

Fig. 4

Once a template is represented by a set of statistics, the task of object detection becomes a straightforward problem in hypothesis testing based on ANOVA. As shown in Figure 5, an image array is transformed and the resultant array is analyzed by ANOVA. If the F- and t-statistics qualify as belonging to a standard pattern of confidence coefficient α, the image data is said to be visually equivalent to the template.

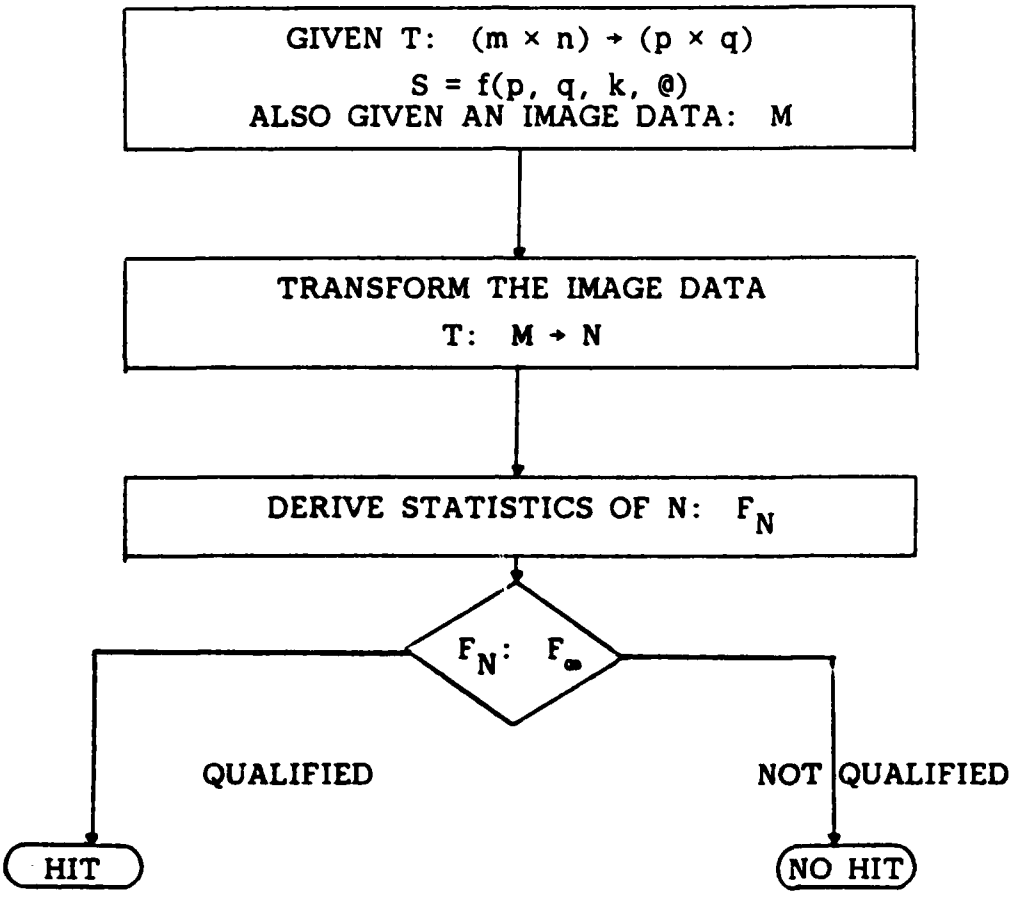


Fig. 5. Object detection scheme.

It is proposed to continue research in this area stressing the following items: study the convergence properties of the proposed object detection scheme, extend the procedure to multiframe object detection, develop a coordinate-free transformation rules generator in the learning scheme. Basic results in this area were submitted for publication.[45]

(2) Trajectory Detection and Experimental Designs

A trajectory is identified by a continuous and narrow region of high intensities which is surrounded by a region of low intensities on each side. When an image containing such a trajectory is processed by a localized operator (window or mask), the grey level distribution within the mask behaves like either a pulse function or a step function,

depending on whether the mask is placed perpendicularly or in parallel to the trajectory. The traditional square-shaped edge detector is ineffective in this application regardless of the relative position of the mask to the trajectory. A small edge detector may not contain enough variations due to trajectory elements to cause significant block effects. On the other hand, a large edge detector may cover too wide an area so that adjacent trajectory segments are included in the same mask. What is needed is an experimental design in the shape of a stretched rectangle. When it is placed perpendicularly, it allows for the inclusion of three component parts of a pulse function. When it is placed in parallel, it allows for the inclusion of substantial elements from only one trajectory segment to form a step function.

A trajectory is further characterized by two inherent properties: the block nature and strong contrast of trajectory elements. Firstly, those elements containing a trajectory tend to be adjacent to one another. Their grey level or texture tend to be similar when compared with the surrounding elements on both sides. Secondly, the grey level distribution changes drastically around both edges of the trajectory. So only those block effects which are associated with strong contrast can be considered as indicative of a "true" trajectory. Based on these properties, an image data can be considered as a three-dimensional array with the row and column numbers describing the planar position, and the grey level distribution is used as the third dimension to describe the relative contrast. In terms of experimental designs, the analysis of an image containing a trajectory calls for a model which can determine the presence of row, column and grey level effects independently.

A SBIB design has the shape of a rectangle due to its definition that the number of blocks is greater than the number of units in each block. Its analysis determines if there exist significant block or grey level effects independently. It also derives the adjusted block means to be used as the unbiased estimators of the true block means.[26] In its application to the trajectory detection problem, the hypothesis that a SBIB mask contains a trajectory is accepted if the following three conditions are all satisfied:

(a) There exist significant block effects.
(b) There exist significant grey level effects.
(c) The adjusted block means behave like a pulse function.

The presence of block effects indicates that there exist edges. When the block effects are qualified by grey level effects, only those edges of strong contrast are preserved. The study of the adjusted block mean indicates that it differentiates a trajectory from an edge.

To satisfy the randomization criterion, a symmetrical balanced incomplete block (SBIB) design is selected from a set of existing designs randomly and then its grey level designations are further randomized by permutations of the row and column assignments. The first two conditions can be easily implemented by comparing the F-statistics of a SBIB design with a threshold value corresponding to certain confidence levels. The shape of the grey level distribution can be determined by thresholding the adjusted block means by their overall means. Further work in this area will be outlined in the section on Proposed Research Program.

(3) Edge Detection Using Radial F-Statistics

Detecting edges separating grey levels is an important aspect of image processing. Here a procedure for detecting edges based on radial F-statistics is proposed. Though the procedure is uniformly most powerful (UMP) for gaussian noise background,[25] it remains robust (insensitive to changes in noise distribution) in many situations.[26] If conditions of reference 26 are not satisfied, robustness of the procedure may be achieved by robustizing algorithms of reference 7. The actual detection method is based on using F-statistics along radial vectors from the inside of the object enclosed by the edge. It can be shown that the radial F-statistic varies linearly between two constant values in passing between two grey levels. The midpoint of the linear portion represents a true edge point. The F-statistic is of the form

$$F_{i;m} = \frac{V_i^2}{V_t^2} = \left[\sum_{i=1}^{m} (S_i - \Sigma\, S_i/m)^2/m\right] \div \left[\sum_{i=1}^{m} (S_{ti} - \Sigma\, S_{ti}/m)/m\right]$$

where V_i^2 = sample variance of the extracted sample; V_t^2 = sample variance of the training sample, and the notation of reference 8 has been used. A paper on this subject has been submitted for publication.[46]

(4) Image Enhancement

The theory of latin squares which was applied to edge detection by Kadar and Kurz[6] can be also successfully applied to image enhancement. Presentation of the methodology is meaningless without demonstration on actual images. Since the proper images are available only on film, the description of this method will be postponed to the oral presentation.

B. Methods Based on Partition and Related Statistical Tests

(1) Classifiers or Detectors Based on Occupancy Vectors

In this section, the theory of m-interval detectors[9] is extended to two-sample detectors based on occupancy vectors. The objective is to find simple, robust detectors for signals in noise with unknown or poorly defined statistics. The Taylor series expansion of the likelihood ratio based on the occupancy vector distributions is used to derive linear and quadratic two-sample tests which are locally most powerful. The first term of the expansion leads to a linear test which is locally more powerful if the relationship of the distributions under the hypothesis (H_0) and the alternative (H_1) is known, i.e., shift of the mean, stochastic ordering, etc. If the relationship is now known, the second term of the expansion leads to a quadratic test which is sensitive to any differences between the received and the reference occupancy vectors. Detectors based on occupancy vectors require specification of m constants $\{a_1, a_2, \ldots, a_m\}$ to partition the space of the distribution $F(x)$ under H_0. Most frequently these constants are chosen to be quantiles $a_i = F^{-1}(1/m)$ which can be easily estimated by methods introduced by the senior investigator.[11,20] Choosing $\{a_i\}$ not coinciding

with quantiles may be useful in some applications especially if the nature of the relationship between H_0 in H_1 is unknown. If $\{x_i\}$ represents the received data (sample) and $\{y_k\}$ the reference or normalization sample, the components of the occupancy vectors for disjoint intervals $A_i = (a_{i-1}, a_i]$ are defined by

$$h_i = \frac{1}{L} \sum_{i=1}^{L} I_{A_i}(x_i)$$

for the received sample and

$$\eta_i = \frac{1}{K} \sum_{i=1}^{K} I_{A_i}(x_k)$$

for the normalization sample. The occupancy vectors for the received (test) and normalization (reference) samples are then $\underline{h}=(h_1, h_2, \ldots, h_m)$ and $\underline{\eta} = (\eta_1, \eta_2, \ldots, \eta_m)$, respectively. It can be shown that the locally most powerful two-sample linear m-interval detector (TLMID) is of the form

$$T = \sum_{i=1}^{m} b_i(h_i - \eta_i) \qquad \text{(B.1)}$$

where the weights (scores) $\{b_i\}$ are selected subject to some optimization criterion. It is proposed to use the Generalized Signal-to-Noise Ratio (GSNR). The GSNR for a test T is defined as $GSNR(T) = \{E[T/H_1] - E[T/H_0]\}^2/var(T/H_0)$ which, for small signal case ($\theta \to 0$), reduces to

$$\lim_{\theta \to 0} GSNR(T) = \left[\frac{d}{d\theta} E[T/H_1]\Big|_{\theta \to 0}\right]^2 \Big/ var(T/H_0)$$

The small S/N GSNR is known as the efficacy. The TLMID gains its robustness from the relative insensitivity to changes in F(x). Crucial to the determination of $\{b_i\}$ is the relationship of F(x) to the distribution under the alternative, G(x), i.e., $G(x) = F(x-\theta)$, $\theta \geq 0$ (shift-of-mean); $G(x) = F(x/(1+\theta))$, $\theta \geq 0$) (change-of-scale); $G(x) = F^{1+\theta}(x)$, $\theta > 0$ (Lehmann alternative), etc. It can be shown, after considerable mathematical manipulations, that for the TLMID the optimum choice of b_i is $b_i = p_i(0)$ where

$$p_i(0) = \frac{d}{d\theta} [F_\theta(a_i) - F_\theta(a_{i-1})]\Big|_{\theta=0}$$

and the corresponding optimum GSNR(T) is

$$GSNR(T) = (LK/L + K)\, m \sum_{i=1}^{m} [p_i(0)]^2 \qquad (B.2)$$

In a straightforward manner one can show that the statistic of (B.1) is asymptotically gaussian. Thus, one can compare TLMID with other detectors based on the asymptotic relative efficiency (ARE). If the noise is gaussian, ARE of TLMID with respect to the optimum parametric detector (OPD) is

$$m \sum_{i=1}^{m} [f(a_{i-1}) - f(a_i)]^2$$

which is the same as for the one-sample m-interval detector.[9] Thus, for m = 6 TLMID reaches over 90% efficiency of the OPD if the noise is gaussian and remains efficient even if burst noise occurs, or the TLMID is robust. In spectral estimation problems the interest is concentrated on the change-of-scale alternatives.[27] Particularly, $g(x) = 1/2^B\, \Gamma(B) x^{B-1} e^{x/2} u(x)$ which is the gamma distribution. Following Woinsky,[27] the efficacy of the OPD is $\varepsilon(t_2) = (LK/L+K)B$, which, compared to the efficacy of the TLMID, yields

$$ARE_{T/t_2} = m/B \sum_{i=1}^{m} [a_{i-1}\, f(a_{i-1}) - a_i f(a_i)]^2 \quad ;$$

for $B \geq 1$ and $m \geq 5$ ARE rapidly approaches one. Thus, the robustness, simplicity of implementation, and efficiency make the TLMID particularly attractive in spectral estimation problems.[27]

The two-sample LMID provides robustness in two forms: by the two-sample detection and by partitioning. The two samples provide robustness against changes in the distribution of the hypothesis of a type similar to the alternative. For instance, in the shift alternative case, the two-sample test maintains performance even if the mean of the background drifts away from zero, where a one-sample test would fail. The second type of robustness is common to partition tests and is due to the unrefined form of partitioning of the space.

In problems where the relationship between F(x) and G(x) is unknown, i.e., poorly defined dispersive media, the two-sample LMID is replaced by a two-sample quadratic m-interval detector (TQMID) which corresponds to the second term of the Taylor's series expansion. A practical form of the TQMID is

$$T_q = \sum_{i-1}^{m} c_i (h_i - \eta_i)^2 \qquad (B.3)$$

If $c_i = 1/p_i$, where p_i is the probability of x_i falling into a given occupancy cell, T_q of (B.3) reduces to the well-known goodness-of-fit test. To find the best weight vector $\{c_i\}$, the approach based on GSNR is not useful. The maximization of the GSNR requires knowledge of the form of the alternative and in the end yields a test sensitive to only that form of alternative. The potential applications of the quadratic test is to detection problems where this knowledge is not available. The weights $\{c_i\}$ are chosen to minimize the variance of T_q of (B.3) under the hypothesis. This approach is related to the one suggested in reference 11. It can be shown that for quantile partitions the optimum $c_i = -(1+2a)/2kb$ where $a = (1/m)(1-1/m)$, $b = (2/m^4 - 3/m^3 + 1/m)$ and

$$k = \sum_{i=1}^{m} c_i \ .$$

Choosing $c_i = 1/m$ makes the TQMID equivalent to a chi-square test with quantile partitions. For partitions other than quantile partitions, as would result from distributions other than the design hypothesis, the optimum weights are not equal. In the latter case, a mathematical relationship between the partition probabilities to the best weights can be used to adapt the detector to provide best performance. For example, a TQMID using fixed partitions in a slowly changing environment might use a periodic learning sample to update the weights to maintain minimum variance operation. This area requires additional work to find a general class of adaptive detectors.

By changing from invariable to bivariate (two-dimensional) statistics, one doubles the effective signal-to-noise ratio.[11] This is of particular importance in processing severely corrupted images. Therefore, the extension of the theory presented in this section to bivariate statistics is important. The extension presents some mathematical difficulties, but the senior investigator is confident that these difficulties can be overcome.

(2) The m-Interval Polynomial Approximation (MIPA) Method

Consider a function $g(x)$ which may represent a p.d.f., a c.d.f., an influence or scoring function in a robust detector, a discriminant function in pattern recognition, etc. The problem posed is:

Given a set of observables (samples) from a population corresponding to $g(x)$, find a robust, adaptive and efficient estimator, $\hat{g}(x)$, of $g(x)$. Consider $\hat{g}(x)$ of the form

$$\hat{g}(x) = \sum_{j=1}^{m} (c_{2j}\, x^2 + c_{ij} x + c_{0j})\, I_{A_j}(x) \qquad \text{(B.4)}$$

where $I_{A_j}(\cdot)$ is the indicator function of the set $A_j = (a_{j-1}, a_j]$ and a_j are found from

$$\int_{a_{i-1}}^{a_j} g(x)dx = \lambda_j, \qquad \sum_{j=1}^{m} \lambda_j = \int_{-\infty}^{\infty} g(x)dx = \text{constant}, \quad a_0 = -\infty, \quad a_m = +\infty$$

such that some convenient measure of error is minimized, i.e., the integral square error (ISE)

$$I(c) = \int_{-\infty}^{\infty} [\sum_{j=1}^{m} (c_{2j}\, x^2 + c_{ij}x + c_{0j})\, I_{A_j}(x) - g(x)]^2\, dx \qquad (B.5)$$

is minimized with respect to $\underline{C}_j = (c_{2j}, c_{ij}, c_{0j})^t$. It can be shown that the optimum constants (scores) can be found from

$$\underline{c}_j = \underline{u}_j^{-1} \underline{\eta}_j$$

where

$$\underline{u}_j = \begin{bmatrix} u_{4j} & u_{3j} & u_{2j} \\ u_{3j} & u_{2j} & u_{1j} \\ u_{2j} & u_{1j} & u_{0j} \end{bmatrix} \qquad u_{kj} = \frac{1}{k+1}\,(a_j^{k+1} - a_{j-1}^{k+1}); \quad \frac{1}{k+1} \quad k = 0,1,2,3,4$$

$$\underline{\eta}_j = [\eta_{2j}, \eta_{1j}, \eta_{0j}]^t \qquad \eta_{ij} = \int_{a_{j-1}}^{a_j} x^i\, g(x)dx, \qquad i = 0,1,2$$

a_j's can be recursively estimated as in references 11 and 20 and η_{ij} can be estimated from

$$\omega_2^j(\ell+1) = \omega_2^j(\ell) + \frac{1}{\ell}\,[[x_\ell]_j^2 - \omega_2^j(\ell)]$$

$$\omega_1^j(\ell+1) = \omega_1^j(\ell) + \frac{1}{\ell}\,[[x_\ell]_j^1 - \omega_1^j(\ell)]$$

$$\omega_0^j(\ell+1) = \omega_0^j(\ell) + \frac{1}{\ell}\,[u(a_j - x_\ell) - u(a_{j-1} - x_\ell) - \omega_0^j(\ell)]$$

where $[x]_j = \begin{cases} x & \text{if } x \in A_j \\ 0 & \text{otherwise} \end{cases}$ and $u(\cdot)$ is the unit step function

All the estimators converge in m.s. sense and with probability 1. In an actual system all the parameters can be estimated simultaneously using a microprocessor operating in a parallel mode. Thus the procedure is continuously adaptive and can be used in a real-time system. It should be noted that the MIPA process outlined above can be modified

by using other than quadratic polynomials in each interval or using different order polynomials in some intervals. The efficacy, robustness, and flexibility of this approach have already been demonstrated in other applications.[16,17] An alternate approach to this problem has been initiated. The new approach uses a smaller number of intervals but a better approximation per interval. The resulting modified techniques permit faster processing of data (smaller sample size needed for the same accuracy as in MIPA).

C. Image Reconstruction Algorithms

In a recent paper[18] the senior investigator introduced recursive algorithms which are useful in image restoration. Unlike methods suggested in the past, which result in ill-posed mathematical problems, the new algorithms converge in the mean square sense and with probability 1. In addition, the algorithms can be easily robustized (see Eq. (14) of reference 18). The immediate goal of future research in this area is extension of the algorithms to two dimensions involving Walsh functions in the two-dimensional reconstruction process. In addition, recent research by the principal investigators suggests that one can obtain a method of robustizing algorithms that is simpler than and as effective as the method described in reference 18. Thus, in extending the procedure to two dimensions the new robustizing process will be included. The methods will be refined to perform more efficiently in specific applications.

A parallel effort will be directed toward developing reconstruction algorithms based on the F-statistic. Consider the problem of edge reconstruction,[28,29] where most of the object was extracted by any suitable method such as the one presented in reference 8. Part of the edge which separates the object from the background is badly corrupted and not recognizable using standard masking techniques. One can think of the corruption as a localized burst noise. The radial F-statistic of Section A-3 can be modified to extract the desired edge. Let the image be distorted locally with the rest of the image noise-free. This renders the application of the standard F-statistic useless because the image elements are of zero variance and known mean. However, if a low-amplitude perturbation pattern, called dither, is added to the picture, the procedure could be applied taking the advantage of the robustness of the F-statistic. Consider a potential edge point S_n. The size of the mask centered on S_n is chosen to preserve the robustness of the statistic and to be able to follow the rate of change in the direction of edges to be classified. Starting with S_n, quantized directional vectors are chosen, and for each quantized directional vector, the gradient is computed at different points. Thus, the mask (image submatrix) is transformed into a differential submatrix of gradients falling below a preset threshold. The steps are as follows: gradients are computed for a number of points in an image containing no object but only background noise; based on the computed values, a threshold is chosen and this threshold is used to generate a mask. This mask is applied to the differential submatrix, and the point S_n is classified as belonging to an edge or not. If S_n is an edge point, the next point S_{n+1} is detected in the direction containing the largest gradient vector and having a location close to the most likely edge direction. Once the

point S_{n+1} is detected, the point lying on the periphery of the sub-matrix is chosen as the next edge element to which the procedure is applied. This is repeated until a closed curve is obtained. If a closed curve does not appear, the procedure is repeated in the opposite direction and interpolation is used to extrapolate the edge. The new application of masks and differential operators can be used not only to reconstruct edges of a partially degraded image but in following the path of a moving object in a noisy environment and in related problems. Simultaneous use of object detection and gradient masks represents another fruitful area for further research. Comparing the relative operational advantages and disadvantages of the two reconstruction philosophies presented in this section will be considered.

D. Robust Recursive Estimators

In this section, a set of computationally attractive Robbins-Monro stochastic approximation (RMSA) procedures, based on the MIPA suggested in the previous section, are introduced. The estimators are asymptotically consistent, efficient, and robust. Extensions to other recursive algorithms are proposed. As a first task, it will be shown that there exists a RMSA equivalent to Huber's M-estimate.[30] Let $x_1, x_2, \ldots, x_n$ be i.i.d.r.v. and from a p.d.f. $f(x-\theta)$. Huber tried to estimate θ by either finding a minimum of

$$\sum_{i=1}^{n} \zeta(x_i - T_n)$$

or by implicitly solving

$$\sum_{i=1}^{n} \Psi(x_i - T_n) = 0 \qquad \text{(D.1)}$$

where ζ is an arbitrary well-behaved function, $\psi(x-\theta) = \frac{\partial}{\partial\theta}\ \zeta(x-\theta)$.

Solutions to Eq. (D.1) yield M-estimates. Under certain regularity conditions, T_n (an estimator of θ) is consistent and $n^{\frac{1}{2}}(T_n - \theta)$ converges in law to the gaussian distribution with mean zero and variance

$$V(\phi,f) = \frac{\int \psi^2\ fdx}{(\int \psi^1\ fdx)^2}$$

For n sufficiently large x,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi(x_i - T_n) = E_\theta[\psi(x-T)]$$

Smoothness of $f(x-\theta)$ ensures the existence of $\partial/\partial T$ and $\partial^2/\partial T^2$ of $f(x-T)$ and their absolute integrability. Let $\alpha(T) = E_\theta[\psi(x-t)]$, then

$$\alpha(T) = 0 \qquad T = \theta$$

$$\neq \text{otherwise}$$

Thus, asymptotically the M-estimator is equivalent to finding the root of the regression function $\alpha(T)$ by using RMSA yielding

$$T_{n+1} = T_n + \frac{1}{n\alpha^1(T_n)} [\psi(x_n - T_n)] \qquad (D.2)$$

where

$$\alpha^1(T_n) = E_\theta \frac{\partial\psi(x-T)}{\partial T} \; T = T_n \quad .$$

Following references 31 and 32, $T_n \to \theta$ in m.s. sense and w.p.l. and $n^{\frac{1}{2}}(T_n - \theta)$ is asymptotically normal of mean zero and variance $V(\psi,f) = \int \psi^2 \, fdx/[\int \psi^1 \, fdx]^2$, which is the same as Huber's result. If $\psi = f^1/f$, T_n is the maximum likelihood estimator (MLE) of θ and has the minimum variance

$$V(-f^1/f,f) = 1/\int(f^1/f)^2 \, fdx = \frac{1}{I_f}$$

where I_f is the Fisher Information of f. In robust estimation theory, f and f^1/f are not known. To circumvent this difficulty, Huber assumes that $f(x-\theta)$ belongs to

$$\Phi_s = \{f|f = (1-\varepsilon)g + \varepsilon h, \qquad h \text{ symmetric}\} \quad .$$

He obtains a min-max estimator by finding ψ which minimizes $\sup\{V(\psi,f):f \in \Phi_s\}$. It should be noted that without symmetry of h, the estimator is biased. Though the theory of min-max estimators is elegant and consistent, it is of limited practical use. Essentially, ambiguity about the knowledge of f has been carried over to the estimation problem of ε and h. Actually, without the knowledge of h, the crucial parameter ε cannot be estimated. It can be shown that wrong assumptions about the h leads to poor estimation of the mixing parameter ε, which is critical in yielding efficient M-estimators. In addition, since this is a nonlinear regression problem, the convergence rate is very slow and accurate estimates are almost never reached for finite sample sizes.

To bypass the difficulties associated with Huber's estimators, the theory of MIPA robust estimators is proposed by this investigator. In the subsequent material all assumptions are relaxed except smoothness of $f(x-\theta)$. Consider a first order MPIA

$$C_1^*(x) = \sum_{j=1}^{m} (C_{1j}x + C_{0j})I_{A_j}(x) \qquad (D.3)$$

where $A_j=(a_{j-1},a_j]m$ h = 1,2,..., j, m;$a_0 = -\infty$, $a_m = +\infty$ and $\int_{a_{j-1}}^{a_j} f(x)dx = \frac{1}{m}$.

The mean-square error E_r^* beweeen $\psi_1^*(x)$ and $-f^1(x)/f(x)$ is

$$E_r^* = \sum_{j=1}^{m}\left[C_{1j}^2\int_{a_{j-1}}^{a_j} x^2f(x)dx + 2C_{1j}C_{0j}\int_{a_{j-1}}^{a_j} xf(x)dx\right.$$

$$+ C_{0j}^2\int_{a_{j-1}}^{a_j} f(x)dx + 2C_{1j}\int_{a_{j-1}}^{a_j} xf^1(x)dx + 2C_{0j}\int_{a_{j-1}}^{a_j} f^1(x)dx$$

$$\left.+\int_{-\infty}^{\infty} [f^1(x)/f(x)]^2f(x)dx\right]$$

The scores C_{1j} and C_{0j} which minimize E_r^* are

$$C_{1j} = \frac{\begin{vmatrix} P_{1j} & \mu_{1j} \\ P_{0j} & a_{0j} \end{vmatrix}}{\begin{vmatrix} \mu_{2j} & \mu_{1j} \\ \mu_{1j} & \mu_{0j} \end{vmatrix}} \qquad (D.4)$$

$$C_{0j} = \frac{\begin{vmatrix} \mu_{2j} & P_{1j} \\ \mu_{1j} & P_{0j} \end{vmatrix}}{\begin{vmatrix} \mu_{2j} & \mu_{ij} \\ \mu_{1j} & \mu_{0j} \end{vmatrix}} \qquad (D.5)$$

where

$$\mu_{2j} = \int_{a_{j-1}}^{a_j} x^2f(x)dx, \qquad \mu_{1j} = \int_{a_{j-1}}^{a_j} xf(x)dx, \qquad \mu_{0j} = \int_{a_{j-1}}^{a_j} f(x)dx$$

$$P_{1j} = \int_{a_{j-1}}^{a_j} f(x)dx + a_{j-1}f(a_{j-1}) - a_jf(a_j), \qquad P_{0j} = [f(a_{j-1}) - f(a_j)]$$

It can be shown that the scores of Eqs. (D.4) and (D.5) also minimize the variance of the estimator with the corresponding

$$V_{min}(\psi_1^*, f) = \sum_{j=1}^{m} (\tilde{C}_{1j} P_{1j} + \tilde{C}_{0j} P_{0j}) .$$

The pertinent parameters can be estimated continuously using a microprocessor operating in a parallel mode based on the recursive procedures presented in Section B.2.

E. Additional Relevant Progress

Some additional recent accomplishments which are relevant to the content of this proposal are included now.

(1) Radar, sonar, and earthsounding images result from particularly weak signals. In addition, the underlying distributions in the latter class of problems are poorly defined. Faced with these difficulties, numerous investigators in this area concentrated their research efforts on array processors, preferably of the robust type.[34-36] Some new results in this area were presented in a paper.[33] The typical assumptions of gaussianity and stationarity of additive noise were removed. In addition, it was assumed that the noise at the sensor inputs are intercorrelated. To robustize the array processing procedure, the optimum nonlinearities at the correlator's inputs were approximated by simplified nonlinear functions based on parameters which can be easily estimated. The effect of the inter-sensor noise dependence on the asymptotic relative efficiency (ARE) of a multiple-input correlation detection procedure was examined. Easy-to-implement detector realizations were shown to be efficient when compared with the optimum detector configurations.

(2) In last year's report three classes of problems -- object reconstruction algorithms,[18] extraction of moving objects by recursive parallelepiped masks[7], and edge detection using Latin square masks[6] -- were described, which depended strongly on a good robust recursive vector parameter estimator. A refined version of such an estimator was generated.[22] An improved rate of convergence and guaranteed low variance of estimation and high degree of robustness (insensitivity to variations in underlying noise distributions) result from the use of rank-type preprocessors. The latter class of preprocessors act as efficient gaussianizers of poorly conditioned data.

(3) The results on dependent sampling in sequential partition detectors reported in reference 12 were extended to a more general model of Markov-dependent noise.[14] This paper involves proving certain basic theorems which point to the fact that considerable advantage in detection efficiency of partition detectors is achieved if one uses the Markov-dependent structure of noise. The results open the door to a whole new class of robust detectors which utilize the structure of dependence of the underlying noise to improve the efficiency of detection.

(4) The advent of modern high-speed sampling techniques resulted in additional stress on the problem of signal processing in dependent noise mentioned in the previous paragraph. Since in many systems of practical interest the knowledge of the noise statistics is either inexact

or unspecified, some adaptive techniques which are frequently ineffective or too complicated to implement were suggested in the past. However, if one applies the methodology of system identification, the approach is essentially equivalent to the operation of a prewhitening filter. In particular, in a recent paper[17] a procedure for robust identification of the autoregressive (AR) model was introduced. The robustizing process is in the form of a robustized Robbins-Monro stochastic approximation (RMSA) algorithm and is based on the m-interval polynomial approximation method. The resulting algorithm represents a recursive robustized version of the well-known maximum entropy method (MEM) for spectral estimation introduced by Burg[37] or of the popular Widrow least-mean-square (LMS) adaptive filter.[38] Furthermore, the robustized algorithm leads naturally to a robustized Akaike's information criterion (AIC).[39] The simplicity of implementation and flexibility make the application of the new robust identification algorithm particularly attractive in practical applications. The flexibility and robustness of the new procedure were confirmed by extensive Monte Carlo simulations. Some of these methods will be tested on real underwater sounding data obtained from the Navy.

(5) The standard Kalman filter will usually diverge when the dynamic noise is not gaussian. A robust Kalman filter based on m-interval polynomial approximation (MIPA) method for nongaussian noise was formulated.[16] Two situations were studied in detail: (a) the states are gaussian and the observation noise is nongaussian, (b) the states are nongaussian and the observation noise is gaussian. It was shown that the MIPA Kalman filter maintains efficient performance over a broad class of noise distributions and is both computationally attractive and easy to implement. Compared to the min-max Kalman filters suggested by other authors, the new class of filters is more efficient and robust. The theoretical results were verified by extensive Monte Carlo simulations.

(6) A particularly efficient and flexible new approach to robust detection was introduced in reference 15. In the paper, the Huber-Tukey model of mixture distributions is replaced by a more versatile one which leads to solutions which are more appealing from a practical point of view than the robust detectors using the Huber-Tukey noise model in conjunction with the min-max criterion of performance. The new detector consists of two parts -- parametric and nonparametric -- which are switched in or out depending on one parameter of the mixture distribution model. As part of the detector, an efficient running estimator of this unknown parameter is included. When compared to the min-max detector proposed by other researchers, the new detector is more efficient for deterministic and stochastic signals. This divergence in efficiencies is particularly apparent under severe and/or variable noise conditions.

(7) The theory of partition detectors was extended to so-called double-threshold detectors which are particularly useful in radar applications.[41] Quantization loss of the commonly used double-threshold detector is minimized by using a multilevel threshold for the first threshold. The overall performance is improved by optimal selection of partitions (thresholds) and scores. Expressions for false alarm rates and detection probabilities and upper bounds based on the generalized Chebyshev inequality are presented. In radar applications, the new detector performs better than detectors suggested by other researchers.

The new detector has potential for application to communications through dispersive media.

(8) The problem of estimating optimum scores for some robust detectors (classifiers) was considered.[42] The detectors of interest are those in the class of partition detectors. They are robust in the sense that a change in the environment will not degrade their performance significantly. The RMSA algorithm was used to obtain recursive estimators. The difficulty with this approach lies in choosing suitable regression functions which are not unique. Two approaches using direct and indirect methods were conceptualized and shown to be sufficiently flexible to yield efficient estimators of scores for almost all partition detectors of practical interest.

(9) The theory of nonlinear and recursive equalizers, presented by the principal investigator in separate papers in the past, was unified to yield nonlinear recursive equalizers which retain good qualities of both classes: immunity to burst noise, and improved efficiency and flexibility.[43] It was shown that it is not always advantageous to reduce the intersymbol interference to zero if the overall mean-square error is to be minimized in the presence of intersymbol interference and noise. The design procedure involved a proper selection of pole-zero locations to ensure optimum performance with minimum sensitivity to selection of parameters in the equalizer. Specific examples were presented in support of the theory developed in the paper.

(10) An orthogonal-mixing concept for adaptive detection in incompletely specified noise was introduced.[44] This is a natural continuation of the robustness principles introduced by the principal investigator through the years. Inherent robustness of some detectors (classifiers) of the LRO or LMIP class was combined with the high efficiency of parametric detectors to obtain good overall performance. A basic theorem on orthogonally combining detectors was proven which leads naturally to the design of numerous practical, efficient, and robust detectors.

(11) The theory of MIPA detection was extended to sequential operation of the detector.[47] The detector operates at near-optimum levels for a particular noise environment and is robust by maintaining high efficiency in other than nominal noise environments by adapting its optimum nonlinearity using an m-interval approximation to it. It was shown that the sequential MIPA detectors are asymptotically optimum, increase their transmission rate up to four times as compared to their fixed-sample-size counterparts, and are highly insensitive to variation of noise if compared to detectors based on min-max theory.

(12) A further improvement in efficiency was shown to be possible in operating robust detectors in a sequential mode. The basic structure of the new class of sequential detectors is based on the theory introduced by Cochrane and Kurz.[49] Though the detector is more complicated than the structure introduced in reference 47, its performance may be improved by as much as 3 db in comparison with MIPA detectors, and the flexibility of design permits inclusion of side constraints in the design procedure.[48]

(13) The concepts of local and global robustness were introduced, which improve our ability to design robust detectors and/or estimators with desirable properties in various regions of operation. Preliminary results on the subject will be presented at the 1983 Symposium on Information Theory.[50]

(14) A recursive receiver structure for receiving Binary Phase Shift Keyed (BPSK) signals over bandlimited, nonlinear satellite channels in the presence of the downlink additive Gaussian noise was derived. The optimum recursive equalizer structure was obtained using the Wiener-Kolmogorov theory. In analyzing the problem, the decision is made on a typical signal in a received sequence taking into account past and future interfering signals, i.e., I.S.I. As an illustrative example of the receiver, a soft limiting channel model was considered in detail. Based on the analysis, an alternative receiver structure which is more suitable for implementation was introduced. The performance of the receiver was evaluated using estensive computer simulation. The new receiver structuare yields significant improvement in performance when compared with a single sample sign detector. Preliminary results pertaining to this satellite communication system were accepted for presentation to the 1983 IEEE Globcom Conference (see reference 52).

5. REFERENCES

1. K.S. Fu and A. Rosenfeld, "Pattern Recognition and Image Processing," IEEE Trans. Comp., Vol. C-2, pp. 358-368 (December 1976).

2. O. Kempthorne, The Design and Analysis of Experiments, Wiley (1952).

3. W.G. Cochran and G.M. Cox, Experimental Designs, Chapter 1, 2nd Ed., Wiley (1957).

4. R.A. Fisher, The Design of Experiments, Hafner, pp. 17-21 (1960).

5. C.A. Mohwinkel and L. Kurz, "Computer Picture Processing and Enhancement by Localized Operations," Computer Graphics and Image Processing, 5, pp. 401-424 (1976).

6. I. Kadar and L. Kurz, "A Class of Robust Edge Detectors Based on Latin Squares," Pattern Recognition, Vol. 11, pp. 329-339 (1974).

7. I. Kadar and L. Kurz, "A Class of Three-Dimensional Recursive Paralleliped Masks," Computer Graphics and Image Processing, 11, pp. 262-280 (1979).

8. S. Kariolis and L. Kurz, "Object Detection and Extraction: A Statistical Approach," Proc. 1980 Conf. Syst. and Info. Sci., Princeton, pp. 500-505 (March 1980).

9. L. Kurz, "Nonparametric Detectors Based on Partition Tests," in Nonparametric Methods in Communications: Selected Topics, P. Papantoni-Kazakes and D. Kazakes, eds., Marcel Dekker (1977).

10. P. Kersten and L. Kurz, "Improved Operation of m-Interval Detectors by Optimum Signal Selection," IEEE Trans. on Info. Theory, Vol. IE-24, No. 4 (July 1978).

11. P. Kersten and L. Kurz, "Bivariate m-Interval Classifiers with Application to Edge Detection," Information and Control, Vol. 34, No. 2, pp. 152-168 (June 1977).

12. R.F. Dwyer and L. Kurz, "Sequential Partition Detectors with Dependent Sampling," J. Cybernetics, Vol. 10, pp. 211-232 (1980).

13. R.F. Dwyer and L. Kurz, "Sequential Partition Detectors in Large Signal and Impulsive Noise," Proc. 1980 Conf. Info. Sci. and Syst., Princeton, pp. 507-512 (March 1980).

14. R.F. Dwyer and L. Kurz, "Theory of Partition Detectors for Data Represented by Markov Chains," IEEE Trans. on Info. Theory (in review).

15. L. Kurz and C. Tsai, "Robust Detection in Mixture Noise," IEEE Trans. on Info. Theory (in review).

16. L. Kurz and C. Tsai, "An Adaptive Robustizing Approach to Kalman Filtering," Automatica, April 1983, pp. 279-288.

17. L. Kurz and C. Tsai, "A Robustized Maximum Entropy Approach to System Identification," IFIP Meeting, New York City (September 1981), also Proceedings of the Meeting, Springer Verlag, pp. 276-284.

18. I. Kadar and L. Kurz, "A Robustized Vector Recursive Stabilizer Algorithm for Image Restoration," Information and Control, 44, pp. 320-338 (1980).

19. J. Evans, P. Kersten and L. Kurz, "Robust Recursive Estimation with Applications," Info. Sci., Vol. 11, No. 1, pp. 69-92 (August 1976).

20. P. Kersten and L. Kurz, "Robustized Vector Robbins-Monro Algorithm with Applications to M-Interval Detection," Info. Sci., Vol. 11, No. 2, pp. 121-140 (October 1976).

21. I. Kadar and L. Kurz, "Robustized Scalar Form of Gladyshev's Theorem with Application to Nonlinear Systems," Proc. 1980 Conf. Info. Sci. and Syst., Princeton, pp. 297-302 (March 1980).

22. I. Kadar and L. Kurz, "A Robustized Vector Recursive Algorithm in Estimation and Image Processing," 1981 IEEE Int. Symp. on Info. Theory, Santa Monica, California (February 1981).

23. A.M. Mood and F.A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill (1962).

24. N.J. Johnson and F.C. Leone, Statistical and Experimental Designs in Engineering and Physical Sciences, 2nd Ed., Vol. II, Chapter 14, Wiley (1977).

25. E.L. Lehmann, Testing Statistical Hypotheses, Wiley (1959).

26. G.E.P. Box and S.L. Anderson, "Permutation Theory in the Deviation of Robust Criteria and Study of Departures from Assumptions," J.R. Stat. Soc., Series B, 17, pp. 1-34 (1955).

27. M.N. Woinsky, "Nonparametric Detection Using Spectral Data," IEEE Trans Info. Theory, Vol. IT-18, pp. 110-118 (1972).

28. H.C. Andrews and B.R. Hunt, Digital Image Restoration, Prentice-Hall (1977).

29. G.T. Herman, "Two Direct Methods for Reconstructing Pictures from Their Projections: A Comparative Study," Computer Graphics and Image Processing, Vol. I, No. 2, pp. 124-145 (1972).

30. P.J. Huber, "Robust Estimation of a Location Parameter," Ann. Math. Stat., Vol. 35 (1964).

31. A. Dvoretzky, "On Stochastic Approximation," 3rd Berkeley Symp., Vol. I, pp. 39-55 (1965).

32. J. Sacks, "Asymptotic Distribution of Stochastic Approximation Procedures," Ann. Math. Stat., Vol. 29, pp. 373-405 (1958).

33. L. Kurz and R.A. Mihajlović, "Robustized Weak Random Signal Array Detection," 1981 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '81) Digest, Vol. II, pp. 858-863 (June 1981).

34. J. Capon, "Application of Detection and Estimation Theory to Large Array Seismology," Proc. IEEE, Vol. 58, pp. 760-770 (1970).

35. G.S. Shin and S.A. Kassam, "Multilevel Coincidence Correlators for Random Signal Detection," IEEE Trans on Info. Theory, Vol. 25, pp. 47-53 (1979).

36. L.C. Wood and S. Treitel, "Seismic Signal Processing," Proc. IEEE, Vol. 63, 1975, pp. 649-661 (1975).

37. J.P. Burg, "Maximum Entropy Spectral Analysis," 37th Annual Meeting Soc. Expror. Geophys., Okalhoma City, Oklahoma (1967).

38. B. Widrow, et al., "Adaptive Antenna Systems," Proc. IEEE, Vol. 55, pp. 2143-59 (1967).

39. M. Akaike, "Fitting Autoregressive Models for Prediction," Ann. Inst. Stat. Math., Vol. 21, pp. 243-247 (1969).

40. L. Kurz and C. Tsai, "Robust Detection in Mixture Noise," IEEE Trans. Info. Theory (in review).

41. S.K. Gambhir and L. Kurz, "Double-Threshold Partition Detectors with Radar Applications," IEEE Trans. on Aerospace & Electr. (in review).

42. L. Kurz and B.P. Rabinowitz, "Estimation of Optimum Scores for Partition Detectors Using Stochastic Approximation Techniques", Proc. 16th Annual Conference on Information Sciences and Systems, March 1982, pp. 598-604.

43. N. Huang and L. Kurz, "An Approach to Nonlinear Recursive Equalization in Data Transmission Systems", Proc. 16th Annual Conference on Information Sciences and Systems, March 1982, pp. 145-150.

44. R.A. Mihajlović and L. Kurz, "Orthogonal-Mixing Adaptive Detection in Incompletely Specified Noise", 1982 IEEE International Conference on Communications Record, Vol. I, pp. 1H-4.1 - 1H-4.6, June 1982.

45. E.S.H. Chang and L. Kurz, "Object Detection and Experimental Designs" (in review).

46. E.S.H. Chang and L. Kurz, "Trajectory Detection and Experimental Designs," Computer vision, Image Processing and Computer Graphics, accepted for publication.

47. E. Voudouri and L. Kurz, "Sequential Robust m-Interval Polynomial Approximation (MIPA) Partition Detectors," Proc. IEEE Int. Conf. on Accoustics, Speech and Signal Processing, April 1983, pp. 611-614.

48. E. Voudouri and L. Kurz, "Generalized Sequential Linear Rank and Partition Tests," Proc. IEEE Int. Conf. on Communications, June 1983, pp. 1597-1603.

49. J. Cochrane and L. Kurz, "Generalized Sign-Test Classifiers within an Unified Framework of M-Sample Nonparametric Tests," Proc. 1972 Princeton Conference on Info. Sciences and Systems, March 1972, pp. 479-483.

50. R.A. Mihajlović and L. Kurz, "On Local Robustness of Some Statistical Tests," to be presented to the IEEE Symp. on Info. Theory, Sept. 1983.

51. C.-L. Lin, "A New Decision Feedback Error Control System with Application in Image Processing," Ph.D. Dissertation, PINY, June 1983.

52. A.F. Elrefaie and L. Kurz, "A Recursive Approach to Detection of BPSK in the Presence of Nonlinearity and Gaussian Noise," accepted for presentation to the 1983 IEEE GLOBCOM Conference.

6. DoD AND OTHER INTERACTIONS

(a) The robust detection techniques developed by Professor Kurz have been of interest to the Navy in solving certain types of underwater sounding problems. His former students were involved with this problem at Bell Laboratories under a contract with the US Navy. In recent years, one of his students (Dr. Robert Dwyer, Underwater Sounding Laboratory, US Navy, New London, CT) worked in collaboration with Professor Kurz on applying statistical partition techniques to underwater sampling data.

(b) The work on statistical masking techniques and image reconstruction algorithms is of interest to the IBM Corp. They have published three internal reports based on Professor Kurz' work in this area, and they hope to include these approaches in their future software packages.

(c) At a recent IFIP Conference, Professor Kurz was approached after his presentation by people from several agencies of the US Government. The Bureau of Census and the Department of Energy, in particular, asked him to give presentations to their staffs on his robustizing approaches to statistical inference.

(d) The concept of the M-interval Partition Approximation (MIPA), particularly in its Kalman filtering form, was selected by the German Academy, Mathematics Division, for publication in an extended abstract form as a "significant contribution to mathematical sciences."

(e) The results pertaining to satellite communications evoked considerable interest form the Comsat Corp. These results are important in improving data transmission through satellite systems.

SECTION IV
SIGNIFICANT SCIENTIFIC ACCOMPLISHMENTS

Polytechnic Institute of New York
Microwave Research Institute

IV. REPORT
ON
SIGNIFICANT SCIENTIFIC ACCOMPLISHMENTS

September 30, 1983

1. A New Absorption Spectroscopy: "Structure Resonance Modulation Spectroscopy"

   Professor S. Arnold
   Work Unit SS3-1

2. An Adaptive Robustizing Approach to Kalman Filtering and Related Problems

   Professor L. Kurz
   Work Unit IE3-2

3. Hybrid Wavefront-Resonance Formulation of Transient Electromagnetic Scattering

   Dr. E. Heyman and Professor L.B. Felsen
   Work Unit EM3-4

4. Rigorous Theory for Obliquely Incident Surface Waves on Dielectric Gratings for Integrated Optics, and its Application to Wavelength Multiplexing in Optical Communications

   Professors S.T. Peng and A.A. Oliner
   Work Unit EM3-1

5. Higher-Order Electromagnetic Resonances of Very Small Particles, and New Physical Implications

   Professor S. Arnold
   Work Unit SS3-1

REPORT ON SIGNIFICANT SCIENTIFIC ACCOMPLISHMENTS

1. A New Absorption Spectroscopy: "Structure Resonance Modulation Spectroscopy"

Professor S. Arnold

A new method has been developed for the direct measurement of the absorption spectrum of single submicroscopic spherical particles, which makes it possible for the first time to determine the absorptive properties of extremely tiny masses, of the order of picograms. The method utilizes the fact that absorption of energy, accompanied by a temperature change, usually results in a change of particle size. Small particles have a rich electromagnetic surface resonance spectrum consisting of extremely sharp spikes, whose location in frequency also depends sensitively on size. By observing the change in visible scattered light near one of the resonances as the particle absorbs energy, say in the IR, we obtain a direct measure of the absorption. The method is sensitive to size changes as small as 10^{-2} Å, and has been shown to yield a fully quantitative mapping of the absorption of particles of known composition and absorption spectrum. This new method, which we have named "Structure Resonance Modulation Spectroscopy," should find applications in the study of many microchemical systems, for example, in the composition and properties of aerosols.

2. An Adaptive Robustizing Approach to Kalman Filtering and Related Problems

Professor L. Kurz

The performance of a linear Kalman filter will degrade when the dynamic noise is not Gaussian. A robust Kalman filter based on the m-interval polynomial approximation (MIPA) method for unknown non-Gaussian noise was introduced. Two situations were considered: (a) the state is Gaussian and the observation noise is non-Gaussian; (b) the state is non-Gaussian and the observation noise is Gaussian. It was shown that, as compared with other non-Gaussian filters, the MIPA Kalman filter is computationally feasible, unbiased, more efficient and robust. Monte Carlo simulations were used to underscore the feasibility, efficiency and robustness of the method. The MIPA Kalman filter performs well and reliably by continuously adapting to the underlying noise conditions using readily available microprocessor technology.

The MIPA methodology has much broader implications than simply robustizing the performance of Kalman filters. If one applies the methodology to system identification, the approach is essentially equivalent to the operation of a prewhitening filter. The robustizing process is in the form of a robustized Robbins-Monro stochastic approximation algorithm. The resulting algorithm represents a recursive robustized version of the well-known maximum entropy method for spectral estimation introduced by Burg or of the popular Widrow least-mean-square adaptive filter. Furthermore, the robustized algorithm leads naturally to a robustized Akaike's information criterion. Other problems in which the MIPA method has found application are estimation of unknown probability density and cumulative distribution functions, discriminant estimation in pattern recognition problems, image reconstruction in unknown noise environment, PCM optimization, design of echo cancelers, robust Viterbi algorithms, etc.

The concept of the M-interval Partition Approximation (MIPA), and particularly its application to Kalman filtering, was selected in September 1983 by the German Academy, Mathematics Division, for publication in an extended abstract form as a "significant contribution to mathematical sciences."

3. Hybrid Wavefront-Resonance Formulation of Transient Electromagnetic Scattering

Dr. E. Heyman and Prof. L.B. Felsen

Transient scattering by a target has conventionally been analyzed either by tracking the impinging wavefront as it strikes the object and interacts with it by multiple traversals of its boundary contour, or by summing over the natural complex resonances (singularity expansion method, SEM). The former approach conveys highly localized information. It is useful at early times when individual wavefront arrivals can be solved at the observer but inconvenient at later times when the memory of previous arrivals blurs the response pattern. The latter approach conveys global information because the resonances encompass the entire object. It is useful at later times when all except a few low-order (low frequency) resonances are highly damped but inconvenient at early times when many (high-frequency) resonances are required to synthesize a zero field on the as yet unexcited portion of the object.

By analyzing the prototype problem of scattering by a circular cylinder, we have shown rigorously that the cumulative effect of multiple wavefront arrivals can synthesize resonances, and the cumulative effect of resonances can synthesize wavefront fields. This has clarified the connection between these distinct wave processes. We have then constructed a hybrid formulation wherein wavefronts and resonances are combined self-consistently within a single unified framework that efficiently describes the response at all times. This new theory is promising for a more systematic analysis of a wide range of transient scattering problems.

4. Rigorous Theory for Obliquely Incident Surface Waves on Dielectric Gratings for Integrated Optics, and its Application to Wavelength Multiplexing in Optical Communications

Professors S.T. Peng and A.A. Oliner

One of the key functions to be performed by integrated optics, in its role as signal processor for optical communications, is that of wavelength multiplexing. A leading component that accomplishes this function involves a succession of gratings placed at a 45° angle with respect to the incoming surface wave, with each grating possessing a different period, the periods so chosen that they reflect the selected wavelengths at the Bragg conditions for each grating. The function is complicated by the fact that there are four stop bands instead of the usual two in the Bragg region because of the oblique angle incidence, and that the theory available in the literature for design is based on a perturbation procedure valid only for shallow grating grooves. Furthermore, it is pointed out in the literature that practical structures require deeper grooves than those for which the available theory is valid.

Our theory for surface waves obliquely incident on grooved dielectric gratings is rigorously phrased, and is valid for rectangular grooves of arbitrary depth. The theory has also provided excellent agreement with accurate measurements taken in West Germany. In addition, it accounts for all of the added complexity with respect to additional stop bands, cross polarization, etc., due to the oblique angle incidence. This theory therefore provides the basis for the design of the integrated optics wavelength multiplexer for optical communications mentioned above.

With this suitability in mind, we have recently calculated curves of the Bragg region stop band widths as a function of various geometrical parameters. It should be noted that our theory is so far the only one from which results can be obtained that are sufficiently accurate and general to serve reliably in component design. These curves, and other related results which should be helpful in the design of these multiplexers (and other components utilizing Bragg reflectors, such as filters), have been presented in June 1983 at the IOOC Conference in Japan.

5. Higher-Order Electromagnetic Resonances of Very Small Particles, and New Physical Implications

Professor S. Arnold

Many studies have concentrated on exploring and exploiting the large local field enhancements on very small spherical particles when the fundamental dipole resonance, occurring at $\varepsilon(\omega) = -2$, is excited. It has so far not been appreciated, however, that, since $\varepsilon(\omega)$ is a very rapidly varying function of frequency ω, there may be an overlapping resonance and therefore a similar field enhancement, in the higher-order spherical modes. We have found, for example, that for Ag particles small enough to be within the traditional Rayleigh limit of dipolar scattering, the electric quadrupole resonance, which we show occurs when $\varepsilon = -3/2$, modifies the internal and external fields significantly in the small frequency domain covering both resonance conditions. The interference of these two leading terms produces an internal field anisotropy, that peaks strongly at the back of the particle. This effect provides the basic source of a reverse anisotropy of energy absorption within the particle that gives rise to a large reverse radiometric response.

There are also some very interesting conceptual implications here. For one thing, the study points to the importance of understanding the distribution of the local field as well as its volume average value, even for very small particles. We have little doubt that non-linear optical interaction will be initiated at the peak fields, and therefore the threshold for such effects cannot be predicted from the volume average field, i.e., Rayleigh theory. Among other potential applications, consequences with respect to the possibility of anisotropic Raman enhancement, and of anisotropic photochemistry, remain to be explored. Rayleigh theory may therefore be applicable most of the time and under most conditions, but the exceptions, which occur at frequencies corresponding to higher-mode electromagnetic resonances, particularly the quadrupole mode resonance, can lead to interesting new physics.

END

FILMED

2-84

DTIC