Engineering Research Annual Report FY 86

UCID--19323-86

DE87 006189

Manuscript date: December 1986

Lawrence Livermore National Laboratory
UCID-19323-86

Foreword

This annual report describes the accomplishments of the Engineering Research Program in the Mechanical Engineering Department for fiscal year 1986. Thirteen projects are covered. For each project, goal(s), technical descriptions, and a financial summary are provided.

Contents

Introduction

Our purpose in conducting engineering research in the Mechanical Engineering Department is to develop mechanical and related technologies to meet the future requirements of Laboratory programs. In fiscal year 1986, the program provided $2 900 000 for 13 projects. The funding consists of Engineering's Institutional Research and Development allocations, weapons research allocations, and that portion of indirect funds previously allocated to exploratory development.

Discipline areas called thrust areas consist of related engineering technologies. We currently have five thrust areas: Computer Code Development, Dynamic Systems and Control, Fabrication Technology, Materials Engineering, and Nondestructive Evaluation. We align projects within these thrust areas to focus our research on meeting the Laboratory's long-term needs.

As part of its mission to stay at the forefront of technology, the Engineering Research Program exchanges information with and transfers technology to technical communities both inside and outside the Laboratory. These extensive interactions are detailed, as appropriate, within each project report.

The intent in this annual report is to provide timely information about research accomplished by members of the Engineering Research Program during fiscal year 1986. Readers desiring more information on the activities reported here are encouraged to contact the individual authors.

J. P. Mahler
Acting Department Head
Mechanical Engineering Department

Solid Geometric Modeling

D. L. Vickers
Weapons Engineering Division

M. K. Kong
Energy Systems Engineering Division

G. W. Laguna and J. R. Kalibjian
Application Systems Division
Computation Department

Solid geometric modeling or "solid modeling" gives a complete, unambiguous analytical description of a solid's geometry and topology. Such modeling will provide the basis for the next generation of computer-aided design (CAD) systems in mechanical engineering. During FY 86 we pursued new avenues of solid modeling research, aiming for payoffs in three major areas: developing a solid modeling link between CAD and computer-aided analysis (CAA), applying solid modeling techniques to actual LLNL components, and developing an understanding of the role and potential of solid modeling within the nuclear weapons complex.

Introduction

Solid geometric modeling is a new and increasingly popular modeling technology that can easily and accurately model three-dimensional (3D) objects. We began our LLNL solid modeling research project in October of 1982. The continuing project goal is to advance the state of the solid modeling art in areas important to LLNL and to pave the way for the eventual widespread use of this technology at the Laboratory. To this end we have already addressed such topics as algorithm research, modeler enhancement, benchmark development, and programmatic utilization. Our aim is not to compete with commercial R&D in terms of modeler development, but rather to concentrate on particular capabilities specific to LLNL needs. We then desire to communicate our innovations to the vendor community for incorporation in commercial products. We view ourselves not as product developers, but rather as inventors, innovators, testers, and publishers.

When our work began, we believed that solid geometric modeling was a fundamentally different approach with significant potential for LLNL. Our experience indicates that solid modeling techniques will lead to improved design methods, fewer design errors, increased involvement of engineers in design, more accurate data bases, significantly increased information on physical and optical properties of parts and assemblies, better interfacing with analysis codes, and better interfacing with manufacturing.

Commercially available solid-modeling systems did not meet all our needs. We undertook our research with the Technical Information Processing System-1 (TIPS-1) and Part and Description Language-2 (PADL-2) solid-modeling codes because they were developed for experimentation and augmentation and because the source codes were available. Also, they take fundamentally different approaches to solid modeling and thus provide useful comparative tests for our innovations.

Summary of Technical Accomplishments

This report summarizes the work of the past year; for a summary of our efforts in the three preceding years, see last year's annual report.[1] We have used two noncommercial, experimental solid modelers for most of our research, namely TIPS-1 and PADL-2. During the past year we have also done some work with a commercial solid modeler, Computervision's Solidesign. Specifically, during FY 86 we accomplished the following:

- Wrote a program that generates a "neutral file" for passing solid modeling data from one vendor system to another. The approach

we used is based on a boundary representation (B-Rep) model and the Experimental Solids Proposal (ESP) neutral file format. We are among the first organizations to implement ESP for a B-Rep modeler. This implementation is a necessary step in eventually communicating complete, "computer-intelligible" product-definition data among sites.

- Modeled a proposed wide-angle refracting telescope using our version of TIPS-1, which is called TIPS-1/LLNL. This version is the only solid modeler in existence that can model both the mass properties and the optical properties of optical components and assemblies.
- Wrote a software tool that gives a finite-element analyst more flexibility in specifying the "lifetime" of an element. This tool was successfully used in analyzing BOA.
- Modeled an unclassified Weapons Engineering Division (WED) part using Solidesign. The results of this project were: (1) accurate mass properties of the part, a highly desirable outcome, since results from the wireframe model were questionable; and (2) familiarity with the strengths and deficiencies of Solidesign.
- Modeled TIPS-1/LLNL to generate uniform rectangular meshes as a first step in creating a finite difference mesh for use in the Electronics Engineering Department's high-power microwave project.
- Managed the nuclear weapons complex (NWC) Advanced Geometric Modeling (AGM) Project, with the goal of working toward compatible implementation of AGM technology throughout the complex.
- Initiated the Advanced Product Modeling (APM) Project within the NWC, aimed at developing a complete product definition model to facilitate automated part analysis and manufacture in the NWC.
- Enhanced TOMAZE to handle more complicated geometries and to be compatible with current engineering practices at LLNL. The TOMAZE program[1] passes two-dimensional (2D) data from our CV CAD system to our 2D mesh generator, MAZE.
- Completed the design of our Television Monitor Display System (TMDS) Frame Grabber. The goal of this effort is to develop a low-cost, high-resolution color display based on existing IBM-PC computers and TMDS cables. The next step is to do the logic simulation.
- Determined the characteristics required of a 3D mesh generator to receive geometric data from a solid modeler. This is a necessary step in the automatic generation of 3D meshes from solid modelers. These characteristics were used to evaluate the suitability of the 3D mesh generator, INGRID, as a receiver for solid geometric data.
- Devised and partially implemented an algorithm to replace the wedge elements generated by the 3D mesh generator, SPIN, with quadrilateral elements. We expect that this algorithm will save WED more than 100 hours of Cray 1-S time each year.

Technical Status

During FY 86 we concentrated our efforts in three avenues of research that appeared to be the most promising and had the greatest relevance to ME Department and LLNL programs (Fig. 1). The first of these was a link to transfer information from LLNL's CAD systems, purchased from Computervision Inc., to our Cray-based analysis codes. Existing links were cumbersome, and there was no path for transferring 3D geometry and topology. Our search led us to Experimental Solids Proposal (ESP), a proposed neutral file standard for transferring solid model data, which we implemented using the Solidesign modeler from Computervision.

We also decided to apply solid geometric modeling techniques to actual design problems at LLNL. For this work we selected typical weapons parts as well as unusual parts that existing commercial modelers could not handle. Our third major effort during the past year was an overview of the potential of solid modeling throughout the NWC. This study included an evaluation of all advanced geometric modeling techniques. The goal was to chart a course for the appropriate implementation of these techniques by the NWC.

CAD to CAA Link

ESP Generator. Through the NWC CIM project, weapons sites are now able to transmit visually equivalent drawings

among themselves. The current medium of exchange is the Initial Graphics Exchange Specification (IGES) version 2.0 data format. Although the IGES format is a significant step in the linking of design and manufacturing systems, it is generally accepted that the limited types of data it is able to transmit will not be sufficient for the future automated manufacturing environment because IGES cannot communicate part topology and associated manufacturing text data (e.g., tolerances, finish specifications, etc.) in a computer-sensible form.

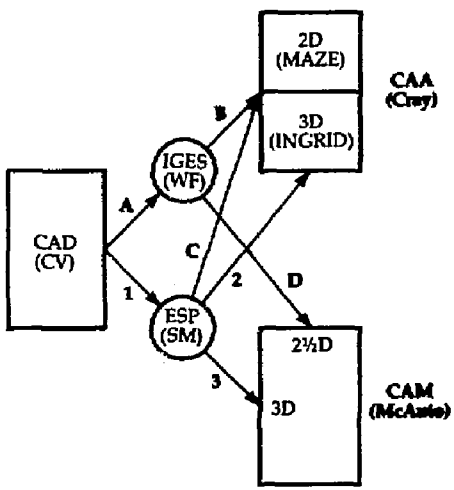
The next step in the evolution of the IGES format is the ability to model part topology. The proposed extension to encompass this is known as the ESP. The proposed ESP standard is able to handle both primary solid model representations: Constructive Solid Geometry (CSG) and Boundary Representation. While there have been many trial implementations of ESP CSG, to this date we know of no one who has yet implemented ESP B-Rep.

We have begun an effort to implement an ESP B-Rep capability on our CV solid modeling system. We feel that this project will have two benefits to the ME Department. First, it will enable us to evaluate the proposed ESP standard and determine whether it meets ME needs. If it is found not to, we will have significant influence to change the standard since we will be one of the few organizations that has actually implemented it. Second, we intend to link an ESP B-Rep capability into one of our 3D mesh generator codes. Thus, the analyst's burden of establishing part geometry and topology in a 3D mesh generator code would be eliminated.

From the technical standpoint, the most difficult part of the project was determining a method by which the ESP file could be generated. We have implemented two differing algorithms to accomplish ESP B-Rep generation. To understand their approach we must quickly review Boundary representation schemes. In a Boundary representation, the relationship of a part's geometrical points, curves, and surfaces is established with the topological notions of vertex, edge, loop, face, and shell. Briefly, a vertex corresponds to a point; an edge to a curve; a closed circuit of edges forms a loop; a face corresponds to a surface; and a closed grouping forms a shell. In a cube, for example, there are 8 points and vertices; 12 curves (lines) and edges; 6 loops; 6 surfaces (planes) and faces; and one shell (see Fig. 2).

The first algorithm traverses CV's data base and establishes 22 cross-reference arrays that relate the geometry and topology of a part. Once this is done, vectorized routines "blast" out the file. Thus, once the cross-reference arrays are established, only simple programming loops need to be executed to produce the file. The second algorithm takes a differing approach: it traverses CV's data base with a post fix traversal of the implicit binary tree established by the B-Rep structure. Thus, whenever a leaf node is encountered, part of the ESP file will be written.

At present, both ESP generator codes have been written on development systems and



- **Bridge footings: IGES, ESP**
- **FY85: A, B, C**
- **FY86: 1, 2, 3, D**

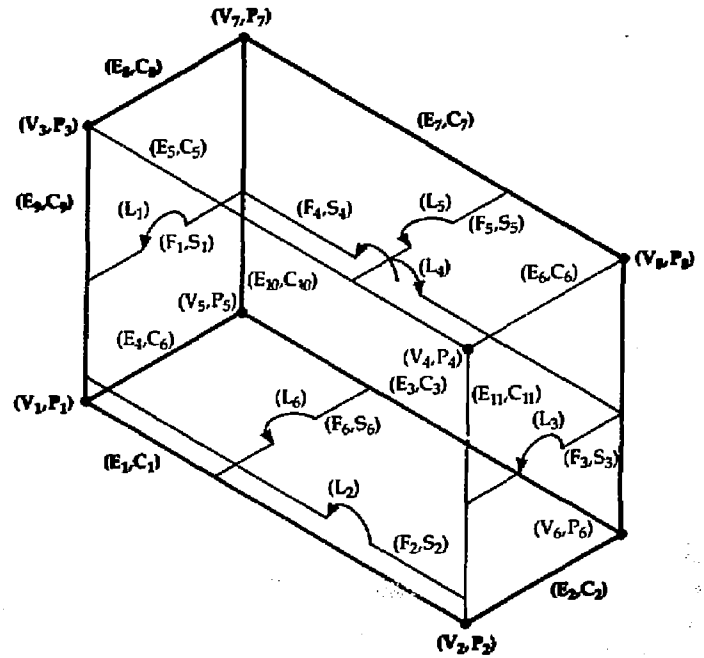
Figure 1. The Mechanical Engineering Department's three "Islands of Excellence," namely CAD, CAA, and CAM, and the bridges connecting them. Proven or installed bridges have alphabetical labels; numbered links are under development. The IGES bridge, currently handling 2D CAA and CAM, operates from a wire-frame data base. The ESP bridge, now being built, operates from a solid-model data base.

work stand alone (with a simulated modeler data base). We are presently porting the code to CV's CAD system and will first try to run them stand alone on the CV's operating system. Once this is accomplished, we will actually try to interface directly to CV's data base.

ESP to INGRID Interface. In tandem with our ESP generator project, we also began an effort to find a 3D mesh generator that could be interfaced with an ESP B-Rep file. We examined the INGRID mesh generator for a month and found that the basic approach of INGRID is not compatible with the ESP representation. The fact that INGRID does not separate the mesh construction activity from the establishment of the geometric/topological part definition prevents it from being easily used in an integrated CAE environment. In INGRID two models must be constructed: (1) an index space model of the mesh, and (2) a geometric definition of the part. These models cannot be constructed separately, but instead must be constructed simultaneously with both spaces "in mind." While this provides the user with "unmatched" 3D meshing capabilities, the requirement to also derive the index space implies automatic mesh generation. Since automatic mesh generation is an unsolved problem, this requirement makes interfacing to INGRID difficult.

To understand this more clearly, consider the two ESP file representations, B-Rep and CSG. Recall that an ESP B-Rep file contains geometric information (points, curves, surfaces) and topological information (vertices, edges, loops, and faces). Although the B-Rep descriptions contain sufficient information to build a geometric definition of

Figure 2. Geometry and topology of a cube.



(V,P) Vertex, point pairs
(E,C) Edge, curve pairs
(F,S) Face surface pairs
(L) Loops

the part, they do not contain any information that can aid in the construction of the index space.

Now consider ESP CSG. In a CSG representation scheme, parts are made up of primitive solids, which are combined using Boolean operators. While INGRID does have 3D primitive geometric entities (no topological information about these entities is maintained), it has no mechanism for evaluating Boolean groups of them. Thus an INGRID real-space definition of the part could not be derived from the ESP CSG description. Actually, index space descriptions may conceptually be easier to derive if one is able to associate a generic index space with each generic primitive. If, then, algorithms could be developed for performing Boolean operations on index spaces, the index space problem would be solved. Unfortunately much theoretical mathematical work would have to be done before an attempt at developing such algorithms could be made. The fact that there are infinitely many real-space-to-index-space mappings introduces additional complexity to the problem.

Thus, we question the viability of linking INGRID with any of the present (IGES version 2.0) or proposed (ESP) neutral file standards. At this point, the best hope for a 3D mesher that can be integrated into a CAE environment using the neutral file transfer concept seems to be one in which geometry and mesh construction are separated, like a 3D version of MAZE. The ESP B-Rep file could be more easily interfaced to such a system.

TOMAZE Enhancements. Extensions are continuing to be added to the TOMAZE preprocessor, which can take 2D IGES geometry and create a MAZE input deck that defines the same part. After an evaluation of TOMAZE using a real part, we have modified the code to enable it to digitize parametric cubic B-splines into MAZE line segments with a completely flexible level of accuracy, and also to extract geometry from one or more IGES layers. We are currently working on an extension that will allow TOMAZE to extract the material property information from each layer of the IGES file. Once this addition is complete, TOMAZE will undergo further tests with actual parts.

Recently LLNL's A Division has become very interested in communicating 2D information between physics and engineering. They have implemented a code that can convert TOMAZE output file geometry into appropriate commands to generate the same geometry in their mesh generator. We are maintaining close communication with A Division to ensure that efforts in the 2D CAD area are not being duplicated.

TMDS Frame Grabber. In FY 86 we designed a device that could plug into an IBM-PC bus and grab TMDS 512 × 512 × 1 pictures. The TMDS data in this frame grabber could then be made available to PC application programs through the PC bus. After a design review in February 1986, the Case CAD system was chosen for schematic input. This phase was completed in June 1986. Our next step was to do a logic simulation to compare with our manual calculations. Yet we could not do this because the system we were on did not have an adequate logic simulation capability. Due to our other commitments, we did not have time to locate a capable logic simulator in the summer months, but have recently begun an effort to find one. Once a timing analysis is done, a wire wrap prototype will be produced and fully tested.

Applications of Advanced Modeling Techniques

Wide-Angle Refracting Telescope. We were requested to contact O Division to ascertain whether our solid modeling resources could help them with one of their SDI projects: the space-based optical telescope. The telescope's most attractive features were an extremely wide-angle viewing capability and low distortion, both of which would facilitate rapid identification and tracking of hostile ballistic missiles. While the SDI office looked favorably on the idea, O Division was worried that it would be difficult to explain the project to Congress, since optical designs are inherently difficult to conceptualize. Thus,

we suggested using TIPS-1/LLNL to model the telescope and then to generate several pictures that would illustrate some of the characteristics of the design. However, before we could generate any images we had to make modifications to the current version of TIPS-1/LLNL.

The telescope consisted of four optical elements, each of which was modeled with several TIPS-1/LLNL primitives. The problem was differentiating between the TIPS-1/LLNL primitives and the optical elements. Without modification, the telescope would appear to be made from 14 optical elements instead of 4. To overcome this problem, the PRODUCT primitive was added to the ray tracing algorithm. The PRODUCT primitive is a valid primitive of TIPS-1, but had never been made functional for the ray tracing program of TIPS-1/LLNL.

This addition to TIPS-1/LLNL required several changes. First, it required changes to the software that calculates the intersection of an arbitrary ray with a surface. Next, it required additions to the algorithm that classifies a point as being inside or outside the solid. Finally, software had to be added to determine the surface normal of a PRODUCT primitive for use in the shading equations.

Once these changes and additions were implemented and tested, several pictures were generated. The first picture (Fig. 3) shows an opaque view of the assembled telescope. The next two pictures (Figs. 4 and 5) show an exploded view of the telescope that illustrates the shape of each optical element. Although the telescope will be manufactured from a transparent refractive material (sapphire), Fig. 4 shows the refractive elements as opaque to depict their arrangement clearly. We then generated images that portrayed the telescope as it would appear when manufactured from sapphire (Fig. 5). Finally, we generated a picture of the assembled telescope showing a simulated ray of light traversing through the lenses and focusing on the back surface of the final lens (Fig. 6).

Having generated the above pictures, we produced one more set to illustrate the angular coverage of the telescope. To illustrate this, we crudely modeled the Earth and a number of observation sites. We then

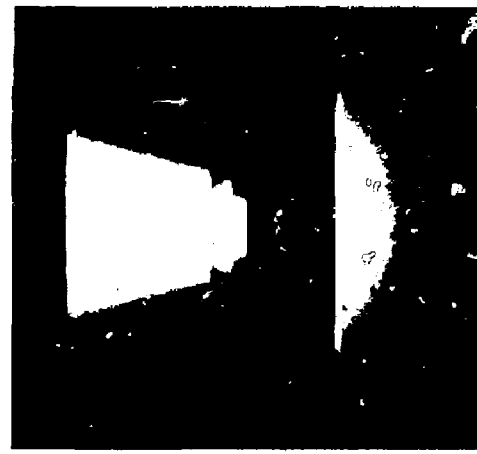


Figure 3. Assembled view of the high-resolution wide-angle optical telescope.

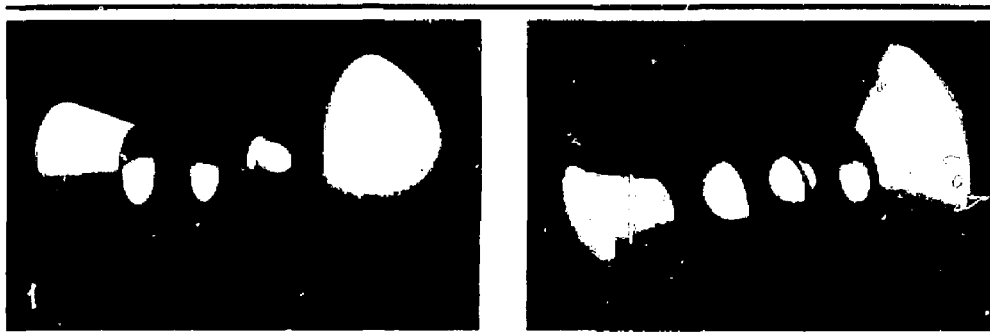


Figure 4. Two views of the telescope with optical elements separated to show shape and function.

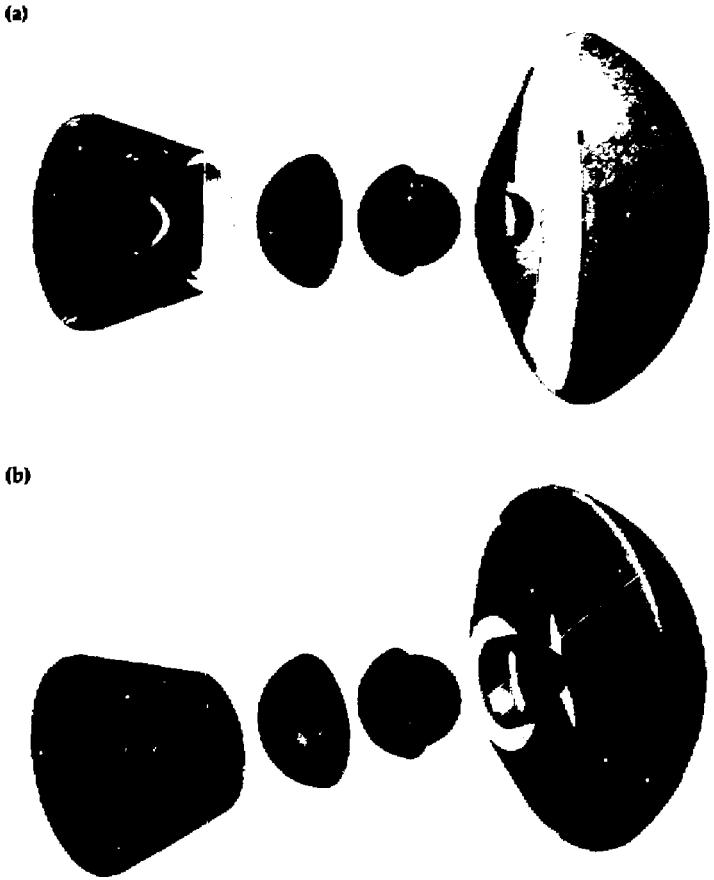


Figure 5. Two exploded views of the telescope showing refraction in sapphire optical elements.

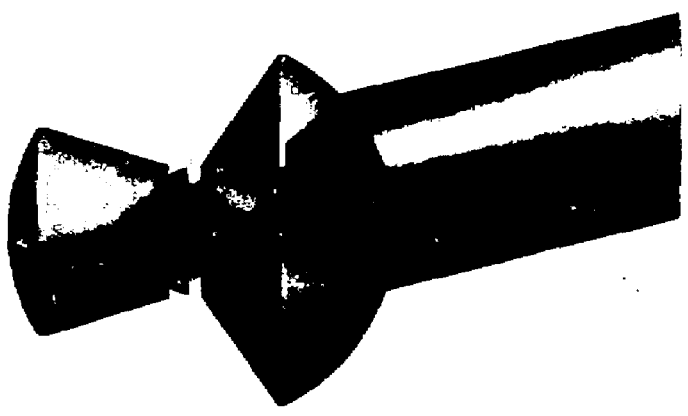


Figure 6. Simulated light beam passing through the telescope optics and focusing on the inner element.

modeled the field of view of the telescope by using a transparent cone with its apex at a location representing the telescope in low Earth orbit. The resulting images are shown in Fig. 7.

WED Part. We were asked to help verify the mass properties for a part of interest within the Weapons Engineering Division (WED). Mass properties are now calculated by manually decomposing a part into simple segments such as cylinders, cones, and toroids. Complex part surfaces are usually approximated by simple quadrics, introducing error into the calculations. We decided to model the part using Solidesign, a solid geometric modeler. Since solid modelers have mathematically complete data bases containing descriptions of all bounding surfaces, mass properties can be computed with a precision equal to the limits of the algorithm and computer used. Thus, the tedious effort of decomposing the part is not necessary. This is not possible with currently used wire frame software because many surfaces are implied and there is no topology to discriminate inside from outside.

We were eager to model this part in order to evaluate the real-world capabilities of Solidesign with actual weapons components. As our modeling proceeded, we became painfully aware of many errors and undocumented limitations. First we noted that subtractions involving two faces that were coincident (one lying on top of the other) could not be performed. We also found that there were limitations concerning the types of objects that could undergo subtraction. Specifically, a solid formed from the union of many primitives could not necessarily be subtracted from a solid formed from a revolved contour. Nevertheless, the part was readily modeled after we became familiar with the Computervision system and the limits of Solidesign. Mass properties calculated using the model were within 1.5% of the values obtained using the cumbersome manual method. We also used the model to generate a shaded image (Fig. 8). This project demonstrated the value of Solidesign in modeling weapons parts and reducing the time required to calculate mass properties.

BOA Mesher. One of our first projects was producing a contour mesh elimination algorithm for the WED analysis group. This algorithm takes as input a DYNA or NIKE input deck and a piecewise (line segment) contour closed curve. The algorithm then marks all mesh elements that lie outside the contour, so that they will quickly die when the mesh is run in an analysis code. Our

Figure 7. Images demonstrating the field of view of the high-resolution wide-angle telescope. The sphere represents the Earth, the cone indicates the field of view of the telescope, and the cylinders represent arbitrarily placed observation sites.

algorithm allows the analyst complete flexibility in selecting an element lifetime.

SPIN Fix. The wedge elements produced in SPIN with the revolution of a 2D part about an axis cost WED approximately 100 hours of Cray time a year in our analysis codes. We have begun an effort to write software that will eliminate those wedge elements and extend the capabilities of the SPIN code. To this point we have accomplished the following milestones:

- SPIN data base restructured.
- Ability to extract arbitrary core from the mesh.
- Ability to graphically display mesh and core.
- Algorithms completed for reconstruction of wedge core for SPIN planes generated with proportional increment ratio of 1.0.
- Ability to display the reconstruction process.

In the next fiscal year, we hope to complete the project by adding the capability to reconstruct the wedge core with spin planes that have been generated with proportional increment ratios other than 1.0; the ability to generate slide planes; and finally the ability to generate the NIKE/DYNA input deck.

Figure 8. Unclassified WED part modeled with Computervision's solid geometry modeler, Solidesign.

Finite Difference Mesh Generation. The EE Department required very large uniform rectangular meshes of weapons systems to perform certain types of electromagnetic analyses for their high-power microwave project. Presently none of the existing mesh generating packages at the Laboratory addressed this need; thus meshes are generated manually. While a former Methods Development Group engineer had started to add this capability to INGRID, his departure from the Laboratory halted all efforts along these lines.

When we learned of this problem, we recognized its similarity to the octree representation of solids. An octree is a tree representation of space occupancy in which space is divided into 8 identical cubes. Each cube is tagged as being empty, partially filled, or completely filled. The process is then repeated for each partially filled cube until all cubes are completely empty, completely filled, or until the size of a partially filled cube is less than the desired resolution. Given an octree representation of a solid, it is easy to subdivide all cubes down to the size of the smallest cube.

We agreed to conduct tests using TIPS-1/LLNL to determine if a finite difference mesh generation application could be driven by a solid modeler. Since our version of TIPS-1/LLNL does not use octrees, we added software to generate a finite difference mesh, using the code's existing capabilities. TIPS-1/LLNL requires that solids be defined with a user-specified domain or box. Our approach was to generate a uniform rectangular grid for this box, where each grid point defines a vertex of a mesh cell. Next the 8 vertices of each mesh cell can be tested to see if the cell is inside or outside the solid. The list of all mesh cells totally inside the solid is sent to TIPS-1 Scanline[1] for graphical manipulation. A sphere was used as a test of this methodology; the resulting mesh is shown in Fig. 9.

As a more advanced test of this algorithm, we generated a simplified model of an F-14A fighter (Fig. 10). The mesh size required to

generate greater detail in this model resulted in files that were too large for our current versions of software. It is clear that the software we developed would require considerable refinement to become a useful tool. However, our project did demonstrate the potential usefulness of solid modeling techniques in mesh generation applications.

Solid Modeling Throughout the Nuclear Weapons Complex

Advanced Geometric Modeling Project. Our leadership in the Advanced Geometric Modeling (AGM) project is closely tied to our engineering research effort in solid geometric modeling. The Advanced Geometric Modeling project is a NWC-wide effort to define and implement the next generation of geometric modeling technology required for NWC applications. The goal of this project has been to work toward NWC-wide compatibility in the use of advanced geometric modelers.

An AGM working committee composed of members from each of the eleven NWC sites has been meeting quarterly for nearly two years. Meetings during the last nine months of FY 85 included extensive in-depth discussions on various aspects of AGM. The committee also conducted a survey within the NWC to explore AGM needs and expectations.

During FY 86 we focused the AGM effort predominantly into three areas, and three task groups were formed within the AGM project to address these areas. The first task group, Implementation, Analysis, and Recommendation, was asked to study various scenarios with single- and multiple-vendor AGM systems within the NWC, and then recommend the most practical and cost-effective approach. The second task group, NWC AGM Requirements Document, was assigned to supplement the FY 85 AGM needs surveys with interviews involving key people at all NWC sites. Their deliverable will be a document summarizing AGM requirements throughout the NWC. The third task group, Vendor Surveys, was given the assignment of comparing the NWC Requirements Document to current vendor offerings in the areas of solid modeling and AGM, and then deciding which vendors best meet the needs of NWC now and in the foreseeable future.

The three AGM task groups are progressing well, with final reports expected in early FY 87. The Implementation, Analysis, and Recommendation group submitted a draft of their report in September of 1986. The report from the NWC AGM Requirements Docu-

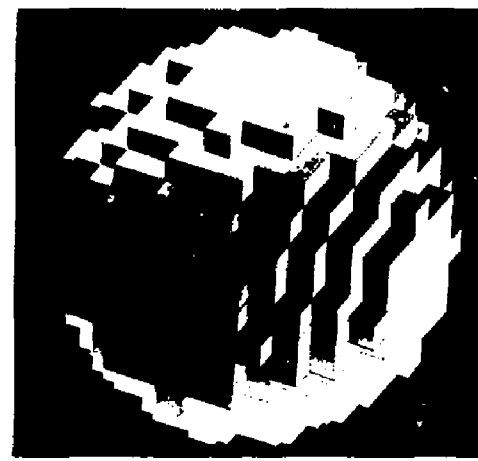


Figure 9. Uniform rectangular mesh of a sphere.

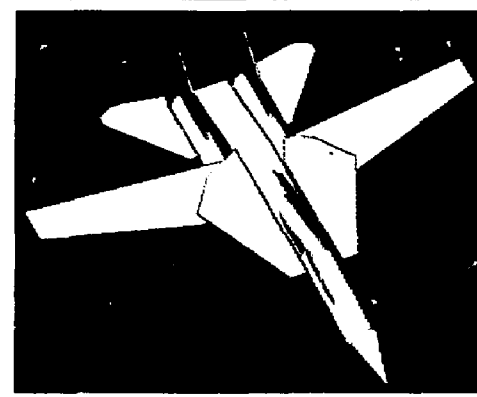


Figure 10. Image of an F-14A from a solid model generated with TIPS-1/LLNL.

ment task group is expected in November of 1986, and the Vendor Surveys report is due in January of 1987.

The AGM project has already received national recognition. We influenced the Computer Aided Manufacturing-International (CAM-I) Geometric Modeling Program to give immediate attention to generating a requirements document. They are even entertaining the idea of making geometric modeling requirements the theme of a nationwide seminar. In addition, several vendors have appropriated the term we invented, and now refer to their future products as "advanced geometric modelers."

Advanced Product Modeling Project. The Advanced Geometric Modeling effort has been redirected to go beyond just the geometry and topology of a product and to include the modeling of a complete product definition. The new initiative, the Advanced Product Modeling (APM) Project, is also NWC-wide. Again, this effort is connected to the Engineering Research Solid Geometric Modeling Project, and we at LLNL have been asked to lead it. The APM is directed toward the exchanging of complete product definition electronically in such a way that it will be able to be used for analysis and manufacturing with less human intervention than is currently required. The stated goal of the APM Project is to (1) develop an advanced product model that contains information for product life cycle support, enabling enhanced exchange and utilization with reduced human interpretation, and (2) provide a path for implementation.

The APM project is in the beginning stages. In July of 1986 a proposal written by LLNL was approved by the Department of Energy Computer-Integrated Manufacturing Program Office. We hosted an APM planning meeting at Livermore on September 16–17, 1986, during which representatives from all eleven of the participating NWC sites agreed on the goals, objectives, and scope of the APM project. At that meeting it was agreed that all NWC activities concerning APM should be closely tied to national product modeling efforts such as the IGES Product Data Exchange Specification project, the Air Force Product Definition Data Interface project, the Air Force Geometric Modeling Applications Interface Program, and CAM-I.

Publications

During the past year we have prepared several papers, including *TIPS-1 User Manual, CRAY-1 Version* (UCID-20545), *Image Realism in Computer Graphics: Theory and Technique* (UCID-20547), and *Extensions to the Simulated Radiograph Capability of the Solid Modeler TIPS-1/LLNL* (UCID-20875).

Future Work

Lawrence Livermore National Laboratory has outstanding facilities for CAD, CAA, and CAM. However, these facilities communicate by the antiquated paper method, which is slow, human intensive, and subject to error. A solid-modeling CAD data base is the key to the eventual automation of CAA and CAM. CAD data bases using 2D line and 3D wire-frame concepts simply do not contain sufficient information to support total automation.

We have made good progress in building a solid-modeling bridge between CAD and CAM-CAA, and we recommend that this work be continued. We have completed the first half of that bridge, namely the ability to generate a neutral file containing a complete solid-modeling data base. The next task is to connect the bridge to our Cray analysis codes, and then to our McAuto manufacturing system. To complete these connections will require three further steps. Step 1 would be to write a neutral file receiver that can understand a complete solid-modeling data base; step 2, to write an interpreter that can turn the neutral file into commands compatible with our 3D mesh generation codes; and step 3, to write a similar interpreter for our CAM system.

Step 1 has been completed, needing only a test with our neutral file generator. Concerning step 2, our investigations during FY 86 suggest that our current 3D mesher INGRID

may not be able to interface efficiently with our solid modeling base because of basic incompatibilities between INGRID and commercial solid-modeling systems. Step 3 cannot be implemented until McAuto adds B-Rep capability to its solid modeler.

With these considerations in mind, our recommendations for future work are as follows:

- Test the link between our neutral file generator and our neutral file receiver.
- Link the neutral file receiver into INGRID. Although this configuration will not pass all the information available in a solid-modeling data base, it will still be a great improvement over current attempts to get 3D CAD data from CV to INGRID.
- Begin a joint ME-EE effort to find or create a 3D mesher that is compatible with solid modelers.
- Continue to exercise all our solid modelers and gain experience with actual LLNL parts.
- Train designers in the use of Solidesign as soon as Revision 5 is released.
- Fabricate and test a TMDS frame grabber prototype.
- Increase our participation in national efforts aimed at automated product definition exchange.
- Support the NWC Advanced Product Modeling Project with LLNL people from the areas most likely to benefit, namely CAD, CAA, and CAM.

Outside Contacts

- In October 1985, M. Kong presented a paper "Implementation of the IGES Experimental Solids Proposal for Data Transfer with Solid Modelers," and also a paper by G. Laguna on our work with simulated radiographs, at the 4th DOE Workshop on CAE, held at Brookhaven National Laboratory.
- In October 1985, D. Vickers gave a talk "Recent Advances in Adapting and Applying Solid Modeling to the Development of Weapons" at the Joint Working Group (JOWOG) 31 Meeting No. 5, held at Atomic Weapons Research Establishment (AWRE) in Aldermaston, England.
- In November 1985, and also in February, May, and September 1986, Vickers chaired the two- and three-day AGM Working Committee meetings held in Los Alamos, NM; Amarillo, TX; Miamisburg, OH; and Livermore, CA, respectively.
- In December 1985, Kong wrote a paper "Integrating Finite Element Analysis with Advanced Geometric Modeling," which was included in the FY 85 Technical Reports of the AGM Working Committee.
- In December 1985, Vickers gave a talk "Solid Modeling and Advanced Geometric Modeling" to the LLNL Director and Associate Directors as part of the 1985 Program Review for Engineering.
- In December 1985 and August 1986, Vickers gave talks to the DOE computer-integrated manufacturing (CIM) Advisory Board describing the status of the AGM Initiative, in Albuquerque, NM, and Golden, CO, respectively.
- In January 1986, Vickers was asked by UCLA to initiate and instruct a short course on Solid Geometric Modeling. The three-day course is scheduled for February 1987.
- In January 1986, Kalibjian gave a presentation "Solid Modeling Foundations for Manufacturing Automation" to the Applications System Division at LLNL.
- In February, May, and July 1986, Vickers gave talks to the DOE CAD/CAM Working Group on the status of the AGM Initiative, in Livermore, CA; Largo, FL; and Albuquerque, NM, respectively.
- In February 1986, Vickers gave an invited talk "Solid Modeling and AGM" at the 3rd Automated Technical Information (ATI) Conference in Livermore, CA.
- In March 1986, Vickers compiled and distributed the five FY 85 Technical Reports written by the AGM Working Committee.
- In March 1986, Vickers gave a talk "The History, Background, and Plans of the AGM Initiative" to the new CIM Program Manager in Albuquerque, NM.
- In April 1986, Vickers gave a six-month Engineering Research progress report to the Laboratory Engineering Executive Committee.
- In May 1986, Kalibjian gave a talk "CAD/CAA/CAM at Lawrence Livermore

National Laboratory and Throughout the Nuclear Weapons Complex" to Dr. Alain de Rosvray of Engineering System International, France.

- In May 1986, Kalibjian gave a talk "CAD/CAA Links at LLNL" to Pete Hearst of AWRE.
- In June 1986, Vickers was invited to speak on "CAE at LLNL" at the engineering portion of the LLNL Summer Lecture Series, Livermore, CA.
- In July 1986, Kalibjian was contacted by Dr. Alan de Pennington, the Program Director of the National Science Foundation, and asked to provide input on solid modeling research at LLNL.
- In July 1986, Vickers gave a talk titled "The Progress in Solid Modeling Research at LLNL" to R Program engineers at LLNL.
- In August 1986, one of G. Laguna's slides of the wide-angle refracting telescope was selected for inclusion in the prestigious SIGGRAPH-86 slide set.
- In August 1986, Vickers and Laguna, along with Peterson of Sandia, Albuquerque, met with the technical staff of Automation Technology Products to discuss strategies for CIM.
- In September 1986, Laguna submitted for publication UCID-20875, "Extensions to the Simulated Radiograph Capability in the Solid Modeler TIPS-1/LLNL."
- Kong has been invited to give a paper "Implementation of an IGES Translator for the MAZE Finite Element Meshing Program" at the October 1986 Computer Use by Engineers (CUBE) Symposium.
- Kalibjian and Laguna have been invited to give a paper "Linking FEM and Solid Modeling Using ESP B-Rep" at CUBE.
- We have been in frequent contact with personnel in Computervision's solid geometric modeling product group. We have used their solid modeler Solidesign as the departure point of our ESP B-Rep generator.
- We have frequently collaborated with the code developers of A Division, who are adapting our TOMAZE for their own finite element analysis needs.
- We have often met with ME analysts in WED and Nuclear Explosives Engineering Division (NEED) to help us adapt and augment TOMAZE to their requirements.
- We have had several exchanges with FMC in San Jose, CA. They are also using CV equipment and are experimenting with Solidesign.
- Our solid modeling research has also brought us into significant contact with other workers in the field, including M. Kmetz, regarding the ESP effort at the University of Wyoming; Dr. L. Rosenblum, from the Naval Research Laboratory and Chairman of the IEEE Technical Committee, concerning an invitation to Laguna to serve on that committee; and Dr. S.-Y. Kung, senior engineer at Applied Research Associates of San Jose, CA, concerning his mesh generator system.

Financial Status

The total expenditure of this project for FY 86 was $192 000 to cover manpower costs and expenses. Computer time was charged to the Weapons Engineering Division Computer-Aided Engineering account.

1. D. L. Vickers, L. E. Taylor, G. W. Laguna, J. R. Kalibjian, M. K. Kong, and F. N. Fritsch, "Solid Geometric Modeling," *Engineering Research Annual Report FY 85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID 19323-85-2 (1985), p. 3.

Investigation of Selected Topics in Structural-Control Technology

S. C. Lu and S. E. Bumpus
Nuclear Test Engineering Division

S. H. Wang
Department of Electrical and Computer Engineering
University of California, Davis

This report summarizes results of integrated structure-control research in three areas: integrated simulation, modeling technique, and decentralized control design. We developed a computer simulation procedure and demonstrated its applicability to the evaluation of controller design with examples. We developed a computer code for the extraction of Lanczos modes. We applied the Lanczos transformation to the solution of classical real and complex eigenvalues and eigenvectors that are commonly used in the normal-mode model reduction method. We derived a theory to adopt Lanczos coordinates as generalized coordinates in the component-mode synthesis technique as an effective modeling scheme for complex structural systems. The theory was illustrated by a simple two-component structural system. Finally, we developed a powerful algorithm for controller design based on an optimal decentralized control theory. In FY 87, we plan to focus on improvements in all three research areas to enhance the applicability of the results to large-scale structural systems.

Introduction

Various types of modern structures, whether they are spacecrafts, precision machine tools, or computer disk drives, often need control systems to assure desired performance. Complex flexible (deformable) structures are usually represented by large mathematical models. The dynamics (vibration) of the structure coupled with its large model size presents a formidable task for control system design. The close interaction between the structural dynamics and the control system means that system and structure must be dealt with in an integrated manner. It has been felt that the state of the art of the technology is inadequate to meet the current and future needs.

In May 1985, an Engineering Research Project entitled *Critical Issues and Required Research in Structural Control Technology* was initiated in the Mechanical Engineering Department, resulting in a research work plan for FY 86.[1]

The FY 86 Engineering Research Project entitled *Investigation of Selected Topics in Structural Control Technology* dealt with integrated simulations, improvements of modeling techniques, and decentralized control design technology.

A computer simulation is essential in evaluating the performance of the control system design before its actual implementation. We developed a computer simulation procedure by linking together a structural analysis code (GEMINI)[2] and a control-design-oriented code (CTRL-C).[3] We exercised the simulation procedure to demonstrate the applicability of the procedure to the evaluation of various control system designs and to identify improvement areas for large complex structural systems.

The modeling technique is an important aspect of the control system design and simulation of large structural systems. Our FY 86 effort concentrates on the exploration of merits of the Lanczos transformation technique[4] with respect to the solution of real and complex eigenvalue problems and the

application to the component-mode synthesis method. The solution of the eigenvalue problem is a crucial step in the normal-mode model reduction scheme, and the component-mode synthesis represents the most effective modeling scheme for complex large structural systems.

To control complex large structural systems, innovative methods for control system design are necessary. We developed a powerful algorithm for control system design based on an optimal decentralized control theory.[5]

Technical Status

Integrated Simulation

Large-scale structures are essentially distributed parameter systems, and their dynamic behavior can be described by partial differential equations. The problem defined by partial differential equations is generally too complex to allow closed-form solution. To circumvent this difficulty, an approach commonly used is to discretize the distributed structure, i.e., to represent it by a discrete model of finite order. The most popular technique of discretization is the finite-element method.[6]

The finite-element model for a structure subject to external forces is usually presented by the following equation of motion:

$$[M]\{\ddot{u}\} + [C]\{\dot{u}\} + [K]\{u\} = \{f\} = \{f_c\} + \{f_d\} \quad , \quad (1)$$

where $[M]$ = mass matrix,
$\{u\}$ = the vector of physical displacements,
$[C]$ = damping matrix,
$[K]$ = stiffness matrix,
$\{f\}$ = vector of external forces,
$\{f_c\}$ = vector of control forces,
$\{f_d\}$ = vector of disturbances,
$\dot{\{\ \}}$ = derivative of a vector $\{\ \}$ with respect to time,
$\ddot{\{\ \}}$ = double derivative of a vector $\{\ \}$ with respect to time.

The control forces $\{f_c\}$ applied to the structure are determined by the control design and usually result in the following expression:

$$\{f_c\} = [C_c]\{\dot{u}\} + [K_c]\{u\} + \{f_r\} \quad , \quad (2)$$

where $[C_c]$ and $[K_c]$ are gain matrices and $\{f_r\}$ vectors are derived from reference input.

Substituting Eq. (2) into Eq. (1) yields

$$[M]\{\ddot{u}\} + [\overline{C}]\{\dot{u}\} + [\overline{K}]\{u\} = \{\overline{f}\} \quad , \quad (3)$$

where

$$[\overline{C}] = [C] - [C_c] \quad ,$$

$$[\overline{K}] = [K] - [K_c] \quad ,$$

$$\{\overline{f}\} = \{f_d\} + \{f_r\} \quad .$$

The simulation of the dynamic behavior is essentially to find the solution for Eq. (3). For complicated structures the dimension of Eq. (3) is usually very large. Normally we need to reduce its dimension by the normal-mode method through the following transformation:

$$\{u\} = [R]\{q\} \quad , \quad (4)$$

where $\{q\}$ is a subset of the normal coordinates and $[R]$ is the mode shape matrix resulting from the eigenvalue problem:

$$[M]\{\ddot{u}\} + [k]\{u\} = \{o\} \quad .$$

After the transformation defined by Eq. (4), Eq. (3) can be expressed as follows:

$$[M^*]\{\ddot{q}\} + [C^*]\{\dot{q}\} + [K^*]\{q\} = \{f^*\} \quad , \quad (5)$$

where

$$[M^*] = [R]^T[M][R] \quad ,$$

$$[C^*] = [R]^T[C][R] \quad ,$$

$$[K^*] = [R]^T[K][R] \quad ,$$

$$\{f^*\} = [R]^T\{f\} \quad .$$

The normal procedure to solve Eq. (5) is to express it in the following first-order form:

$$[E]\ \{\dot{x}\} = [A]\ \{x\} + \{b\} \quad , \qquad (6)$$

where

$$[E] = \begin{bmatrix} I & o \\ C^* & M^* \end{bmatrix} ,$$

$$\{x\} = \begin{Bmatrix} q \\ \dot{q} \end{Bmatrix} ,$$

$$[A] = \begin{bmatrix} o & I \\ -K^* & o \end{bmatrix} ,$$

$$\{b\} = \begin{Bmatrix} o \\ f^* \end{Bmatrix} .$$

We can see now that the simulation of the dynamic system rests on (1) the formulation of the finite-element model expressed by Eq. (1), and (2) the solution of the first-order differential equation, Eq. (6).

In the simulation procedure we developed for this research project, we use the computer code GEMINI[2] to provide the finite-element model and the computer code CTRL-C[3] to provide the solution for Eq. (6).

GEMINI is a computer program developed for the finite-element analysis of linear elastic structures on CDC 7600 or Cray computers at the Lawrence Livermore National Laboratory. On the other hand, CTRL-C is a computer-aided design system that provides commands for matrix analysis, control system design and simulation, and other operations. It operates on a VAX computer. We have developed the necessary interface to link the two codes, which reside in two different computer systems.

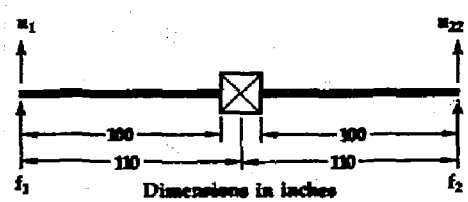


Figure 1. Simplified spacecraft model.

To demonstrate the capability of the simulation procedure, we have applied it to two examples. The first example is a simplified spacecraft model with two flexible appendages derived from Ref. 7 (see Fig. 1). The objective of this simulation is to evaluate the adequacy of the control design for the spacecraft, which is based on the assumption that appendages are rigid.

The finite-element model has 10 beam elements for each of the appendages. The polar moment of inertia about the center of gravity of the spacecraft is 9000 lb·s^2·in. The appendage has the following properties: cross-section area, 440 in.^{2}; area moment of inertia, 1333 in.^{4}, Young's modulus, 10^7 psi; mass density, 2.5×10^{-4} lb·^{2}/in.^{4}. The control design places proportional and derivative (PD) controllers at the tips of the appendages. The control forces are therefore defined by

$$f_1 = -k_p\, u_1 - T_d\, K_p\, \dot{u}_1 + k_p\, r \quad ,$$

$$f_2 = -k_p\, u_{22} - T_d\, K_p\, \dot{u}_{22} - k_p\, r \quad ,$$

where f_1 and f_2 are control forces at the tips, u_1 and u_{22} are transverse displacements at the tips, k_p and T_d are feedback controller parameters, and r is the reference input. According to Ref. 7,

$$K_p = 26.92 \text{ lb/in.} \quad ,$$

$$T_d = 0.06366 \text{ s} \quad .$$

Based on the rigid-body model, these parameters are chosen such that the closed-loop system has 20% damping ratio and 1-Hz natural frequency. In our closed-loop system model, we have included one rigid body and three flexible modes. The response at the tip to a unit step input [$r(t) = 1$ for $t \geq 0$] is shown by Fig. 2. As we can see, the tip of the beam approaches 1 in. as desired. Even though the PD design was based on a rigid-body model, the flexible modes do not affect the system response.

To emphasize the effect due to the flexible modes, we add a point mass of 10 lb at nodes 6 and 16, respectively. We adjust the feedback parameter K_p upward so that the rigid-body system has the same eigenvalues.

The simulation results are in Fig. 3. We can see that structural modes are excited and affect the tip response significantly.

To control a complex large flexible structure, innovative design methods are necessary. In this research, we proposed to apply the Decentralized Optimal Control Theory[5] to our problem. The research effort on decentralized control design will be addressed elsewhere in this report.

To evaluate the effectiveness of the Decentralized Optimal Control Theory, we adopt a specific example from Ref. 8. In this example a 1-m-long flexible beam structure, as shown in Fig. 4, is to be controlled. The beam is pivoted in the middle, and has two 1-kg lumped masses at 30 and 70 cm from one end, respectively. The beam has a Young's modulus $E = 10^9$ (meter)^4, which corresponds to a 1-cm-square cross section. Without the extra masses, the beam has a total mass of 1 kg. The torsional spring is rather weak with $k = 0.1$ N·m/rad and the linear spring is very strong at $k_1 = 10^{10}$ N/m.

Two independent control forces, f_1 and f_2, are applied to each end of the beam. We have derived a finite-element model for this system with 51 modes.

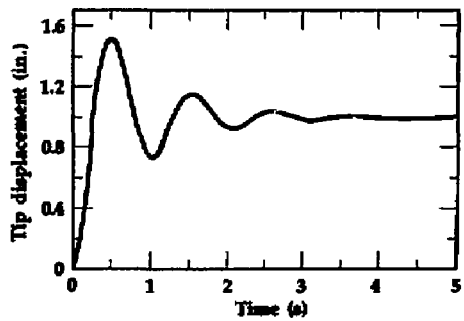


Figure 2. Stable response at the tip of the spacecraft.

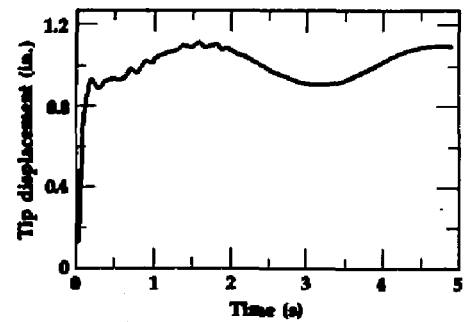


Figure 3. Unstable response at the tip of the spacecraft.

Figure 4. Configuration of the flexible beam.

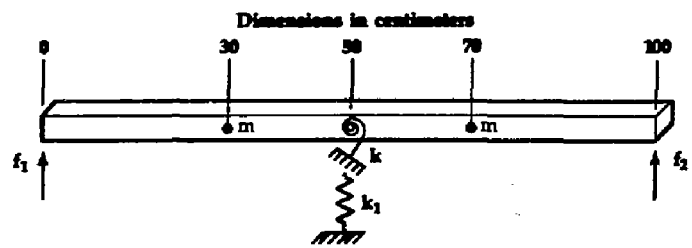


Figure 5. Initial beam deformation.

The Decentralized Optimal Control technique results in feedback law as follows:

$$f_1 = -14.2045 \times (\text{displacement at left tip}) - 3.9060 \times (\text{velocity at left tip}) ,$$
$$f_2 = -14.2045 \times (\text{displacement at right tip}) - 3.9060 \times (\text{velocity at right tip}) .$$

Next we perform the time-domain simulation. For this regulation problem, we simulate the time response of the controlled beam starting from the initial bending condition shown in Fig. 5. The smooth time responses of both tips are shown in Fig. 6.

Modeling Technique

The infinite-dimensional distributed parameter system, which represents a complex structure, can be reduced to a finite-dimensional model by a discretization scheme such as the finite-element method.[6] The dimension associated with the finite-element model frequently has to be very large to preserve sufficient fidelity to the actual structures. This dimension is usually too large for the successful implementation of controllers. A variety of methods may be employed to produce a reduced-order model.

One of the effective model reduction schemes is to find a subspace orthonormal transformation to reduce the equation of motion to a much smaller number of degrees of freedom. One example is the commonly used normal-mode reduction method. In this case, the transformation is the one involving free-vibration mode shapes and frequencies of the undamped structural system.

For structural systems of extreme complexity, different portions of the structure are often handled independently before the final assembly. The component-mode synthesis method[9] has been steadily developed over the last two decades to meet the demand for this independence in handling complex structures. The complex structural system is broken into substructures (or components) for component-mode synthesis. The dynamic characteristics of each component are usually expressed by the free-vibration modes and freqencies associated with the component.

Application of Lanczos Transformation to Eigenvalue Problems. It is recognized that the normal-mode transformation is not the only transformation that can be used to produce a reduced-order model. Recently, Nour-Omid and Clough[4] proposed a procedure to calculate Lanczos coordinates that will serve as the orthonormal dynamic subspace transformation. Because the algorithm for calculating Lanczos coordinates does not necessarily orthogonalize the current vector with all previously calculated "modes," the Lanczos transformation appears to be a more efficient model-reduction algorithm. In our research, we have explored the Lanczos transformation as a model-reduction scheme to serve as an intermediate step for the extraction of approximate eigenvalues and eigenvectors to be used in the model-reduction method. We also explored Lanczos coordinates as generalized coordinates to replace normal coordinates in the component-mode synthesis modeling technique.

For an undamped dynamic system, such as

$$[M]\{\ddot{u}\} + [K]\{u\} = \{f\} , \quad (7)$$

where $[M]$ is the mass matrix, $[K]$ the stiffness matrix, $\{u\}$ the displacement vector, and $\{f\}$ the force vector, the Lanczos transformation,

$$\{u\} = [Q]\{q_L\} , \quad (8)$$

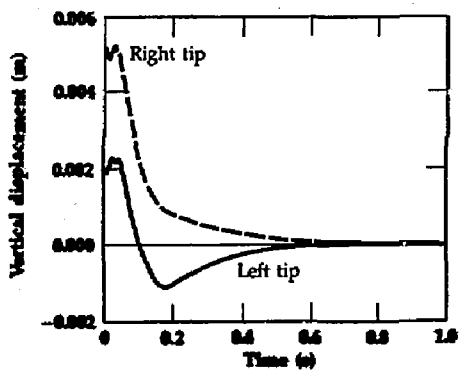


Figure 6. Response at tips of the flexible beam.

possesses the following properties:

$[Q]^T[M][Q] = [I]$ (an identity matrix) ,

$[Q]^T[M][K]^{-1}[M][Q] = [T]$ (a tridiagonal matrix) .

As a result, Eq. (7) will be transformed into the following reduced and simplified form:

$$[T]\{\ddot{q}_L\} + [I]\{q_L\} = \{f_L\} \quad , \tag{9}$$

where $\{f_L\} = [Q]^T[M][K]^{-1}\{f\}$.

From Eq. (9), an approximate subset of the real eigenvalues and eigenvectors of Eq. (7) can be easily extracted by solving the eigenvalue problem $[T]\{\ddot{q}_L\} + [I]\{q\} = \{0\}$.

For a damped dynamic system

$$[M]\{\ddot{u}\} + [C]\{\dot{u}\} + [K]\{u\} = \{f\} \quad , \tag{10}$$

the Lanczos transformation results in

$$[T]\{\ddot{q}_L\} + [C_L]\{\dot{q}_L\} + [I]\{q_L\} = \{f_L\} \quad , \tag{11}$$

where $[C_L] = [Q]^T [M] [K]^{-1} [C] [Q]$.

From Eq. (11), we have the following complex eigenvalue problem

$$[A]\{\dot{x}\} = [B]\{x\} \quad , \tag{12}$$

where $\{x\} = \begin{Bmatrix} q_L \\ \dot{q}_L \end{Bmatrix}$,

$$[A] = \begin{bmatrix} I & 0 \\ C_L & T \end{bmatrix} ,$$

$$[B] = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} .$$

Table 1 shows the real eigenvalues (or frequencies) for the first four modes of a cantilevered beam extracted by the Lanczos transformation technique. Table 2 shows the complex frequencies for a shallow beam clamped at one of the ends. The complex modes were introduced by a single dashpot attached to the structure.

Application of Lanczos Coordinates to the Component-Mode Synthesis. The fundamental theory and the detailed mathematics involved in this research effort are quite lengthy and therefore will be presented in a forthcoming report.[11] In the present report, the component-mode synthesis of a two-component structural system is described to illustrate the procedure. The overall structure is a fixed-fixed beam with unit length and unit properties (bending modulus, area, mass density) and is divided into two components.

Component 1 is a 0.6-unit-long cantilevered beam fixed at the left end. Component 2 is a 0.4-unit-long cantilevered beam fixed at the right end. The finite-element model for Component 1 has 13 nodes and 12 elements, while that of Component 2 has 9 nodes and 8 elements.

The component-mode synthesis of the unit-long beam included the Lanczos coordinates that were generated for the 0.6- and 0.4-long cantilevered beam models as generalized coordinates. Four Lanczos modes and three Lanczos modes were extracted, respectively, for Component 1 and Component 2. In addition, static shapes were also included as part of the generalized coordinates for synthesis. These shapes were obtained by applying a corresponding static load to each degree of freedom at the boundary interface. That is, a unit force in the lateral direction was applied at the translational degree of freedom at the tip of each cantilevered beam. Also a unit moment was

Table 1. Frequencies in Hertz obtained by the Lanczos technique vs the direct solution.

Mode	Lanczos solution	Direct solution
1	9.26002×10^{-4}	9.26002×10^{-4}
2	6.03160×10^{-3}	6.03160×10^{-3}
3	1.75297×10^{-2}	1.75293×10^{-2}
4	3.71372×10^{-2}	3.78727×10^{-2}

Table 2. Complex frequencies in Hertz obtained by the Lanczos technique vs experimental results.

Mode	Lanczos solution	Experiment[10]
1	59.16	52.88
2	370.66	331.57
3	1029.78	935.30

applied at the rotational degree of freedom at the tip of each beam.

With these 11 generalized coordinate shapes—4 Lanczos coordinates and 2 static shapes for the 0.6-unit-long beam plus 3 Lanczos coordinates and 2 more static shapes for the 0.4-unit-long beam—a component-mode synthesis was performed using the method that Flashner[12] described. There are two boundary constraints on this problem, namely the deflection and rotation at the boundary of the two components. The 11 generalized coordinates and the 2 constraints result in 9 unconstrained generalized mathematical coordinates in the formulation. From these nine coordinates, the full nine eigenvalues were extracted. The e values are shown in Table 3 with compara e values obtained by the direct solution o. the eigenvalue problem. The eigen ector results were back substituted in all the generalized coordinate equations to give the synthesized mode shapes for the overall unit beam.

Table 3. Eigenvalues in Hertz obtained by the component-mode synthesis (CMS) vs the direct solution.

Mode	CMS solution	Direct sol ...
1	3.55629	3.55626
2	9.7 6867	9.76862
3	19.0676	19.0459
4	31.4538	31.2495
5	46.6023	46.2437
6	70.9796	63.8619
7	110.281	83.9132
8	142.353	106.186
9	253.670	130.456

To further demonstrate the capability of the component-mode synthesis, we also calculated transfer functions and conducted a forced response simulation.

Overplots of transfer functions using the derived mode shapes and frequencies from component-mode synthesis and from the undamped eigenproblem solution are given in Figs. 7 and 8. The overplots show very little discrepancy in either amplitude or phase until mode number 6 at about 400 radians/s is reached. Forced time-history response corresponding to the transfer functions were generated to show the capability to perform such analyses. The input moment was taken as the north-south component of the 1940 El Centro earthquake because it has

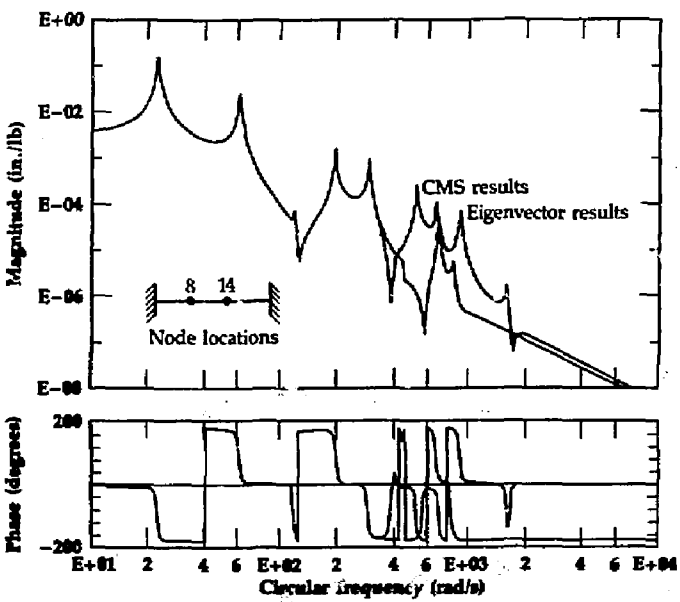


Figure 7. Displacement transfer function at node 14 due to a force input at node 8.

a broad frequency spectrum. The moment was taken as the evenly digitized acceleration, completely unscaled. The time scale was scaled down by a factor of 4 so that all frequency content was shifted lower, which is nearer to the frequency range of the unit beam. The plot for the case where the mode shapes and frequencies are obtained from component-mode synthesis is shown in Fig. 9. The plot for the standard eigenvalue simulation would normally be shown separately, but there is so little difference that the plot in Fig. 9 can serve instead. In fact this figure is the overplot, which shows only a little mismatch that occurs due to some small response of modes at high frequencies.

Decentralized Control System Design

To control complex large structural systems, innovative methods are necessary. In this research, we propose to apply the Decentralized Optimal Control Theory[5] to our problem.

In Ref. 5, Yanchevsky and Hirvonen have proposed an interesting algorithm, which constructs a sequence of feedback matrices for decentralized optimal control problems. Specifically, consider the following discrete-time system:

$$x_{t+1} = Ax_t + Bu_t \quad , \tag{13}$$

where $x_t \in R^n$, $u_t \in R^m$ are the state and input, and A and B are constant matrices of appropriate sizes. The usual quadratic performance index is defined as

$$J = x_T' Q_T x_T + \sum_{t=0}^{T-1} (x_t' Q x_t + u_t' R_0 u_t) \quad , \tag{14}$$

where Q_T and Q are assumed to be positive semidefinite, and R_0 is assumed to be positive definite.

The optimal control problem is to find a sequence of matrices K_t, such that the feedback control $u_t = -K_t x_t$ minimizes the performance index J in Eq. (14). Let k_t^{ij} be the element at the ith row and the jth column of the matrix K_t. Let u_t^i and x_t^i be the ith compo-

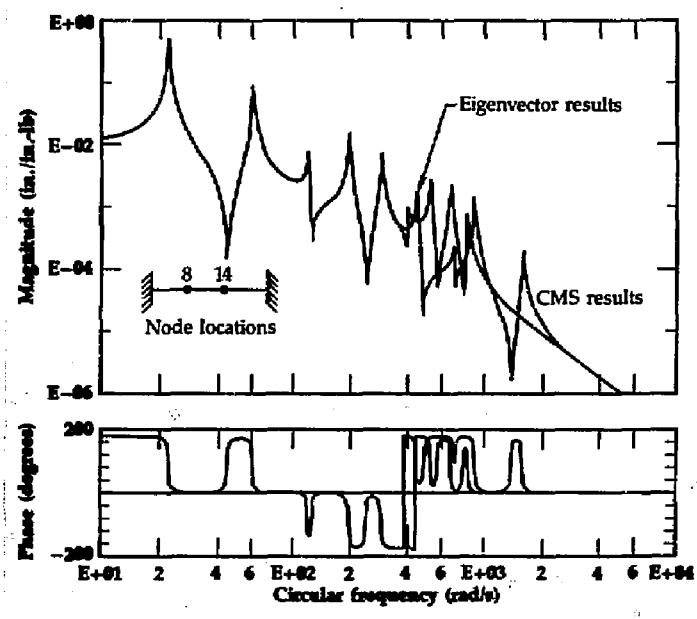


Figure 8. Displacement transfer function at node 8 due to a torque input at node 14.

nent of u_t and x_t, respectively. Then u_t^i can be written as

$$u_t^i = u_t^{i,1} + \ldots + u_t^{i,n} = -k_t^{i,1} x_t^1 - \ldots - k_t^{i,n} x_t^n = -\sum_{j=1}^{n} k_t^{i,j} x_t^j \quad .$$

The augmented performance index is as follows:

$$J = x_T' Q_T x_T + \sum_{t=0}^{T-1} (x_t' Q x_t + u_t' R_0 u_t + \sum_{i=1}^{m} \sum_{j-1}^{n} u_t^{i,j} r_{i,j} u_t^{i,j}) \quad . \tag{15}$$

The nonnegative numbers $r^{i,j}$ may be interpreted as the cost of feedback from the jth component of x_t to the ith component of u_t.

Let e^i denote an n-dimensional column vector whose ith component is equal to 1 and others are equal to 0. Then the performance index shown by Eq. (15) can be rewritten as

$$J = x_T' Q_T x_T + \sum_{t=0}^{T-1} (x_t' (Q + K_t' R_0 K_t + \sum_{i=1}^{n} e_i e_i' K_t' R_i K_t e_i e_i') x_t \quad , \tag{16}$$

where R_i are diagonal matrices $R_i = \text{diag}\,(r_{1,i}, r_{2i}, \ldots, r_{m,i})$, $(i=1, \ldots, n)$. Using dynamic programming techniques to minimize Eq. (16), the following iterative formula is obtained:

$$K_{t-1}^i = (B' P_t B + R_0 + R_i)^{-1} D_t^i \quad , \quad (i = 1, \ldots, n) \quad ,$$

$$P_{t-1} = Q + (A - BK_{t-1})' P_t (A - BK_{t-1}) + K_{t-1}' R_0 K_{t-1} + \sum_{i=1}^{n} e_i e_i' K_{t-1}' R_i K_{t-1} e_i e_i' \quad , \tag{17}$$

$$D_t = B' P_t A \quad ,$$

$$P_T = Q_T \quad ,$$

$$J = x_0' P_0 x_0 \quad ,$$

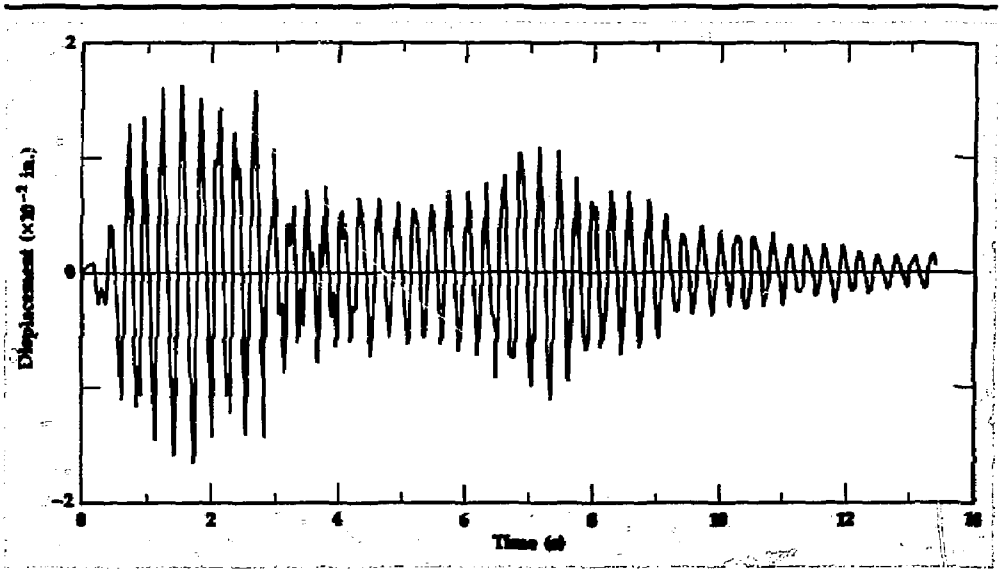


Figure 9. Time-history response for unit beam.

where D_t^i and K_t^i, $(i=1, \ldots, n)$ are the ith columns of matrices D_t and K_t, respectively.

Equation (17) generates a sequence of feedback matrices K_t for the performance index in Eq. (16), and it converges to the steady-state solution K if system Eq. (13) is stabilizable. We consider the case of decentralized control by setting the cost of feedback $r^{i,j}$ to either 0 or $= \infty$, depending on whether the state x^j is available for evaluating u^i or not. Equation (17) can be modified as follows:

$$K_{t-1}^i = F((B'P_tB + R_0), R_i)\, D_t^i \quad ,$$

$$(i = 1, \ldots, n) \quad ,$$

$$P_{t-1} = Q + (A - BK_{t-1})'P_t(A - BK_{t-1}),$$

$$+ K_{t-1}'R_0K_{t-1} \quad , \qquad (18)$$

$$D_t = B'P_tA \quad ,$$

$$P_T = Q_T \quad ,$$

$$J = x_0'P_0\,x_0 \quad ,$$

where $F(\cdot)$ is defined as follows:

$$F((B'P_tB + R_0), R_i)$$

$$= \lim \ldots \lim (B'P_tB + R_0 + R_i)^{-1} \quad ,$$

$$r^{i,j} \to \infty, (i,j) \in S \quad ,$$

and $S = \{(i,j) \mid k_t^{i,j}$ is required to be zero for all $t\}$.

To evaluate the effectiveness of the Decentralized Optimal Control Theory, we adopt a specific example from Ref. 8, which represents a flexible beam as described previously.

To apply the Decentralized Optimal Control Technique, we exclude all but the lowest six modes in the design model. This continuous-time design model is further transformed into a discrete-time system using a sampling time of 0.01 s.

In Eq. (15), the weighting matrices Q_T and Q are chosen to be $50 \times I_{12}$, and $R_0 = 10^{-5} \times I_2$. We further assume that at each tip of the beam there is a displacement sensor and a velocity sensor to produce the information for feedback.

By using the formula in Eq. (17), we obtain the decentralized feedback law as follows:

$f_1 = -14.2045 \times$ (displacement at the left tip) $- 3.9060 \times$ (velocity at the left tip).

$f_2 = -14.2045 \times$ (displacement at the right tip) $- 3.9060 \times$ (velocity at the right tip).

Then we apply this decentralized feedback to the 12th order design model; the resulting closed-loop system is stable with the following eigenvalues:

$0.7625 \pm 0.1933i$,
$-0.4212 \pm 0.7023i$,
$-0.5613 \pm 0.5999i$,
$0.7925 \pm 0.2883i$,
$0.8426 \pm 0.1087i$,
$0.9226 \pm 0.0255i$.

The largest magnitude of these eigenvalues is 0.9230.

Next we perform the time-domain simulation to evaluate the performance of the controller. The smooth time responses of both tips are shown in Fig. 6. For more details, see Ref. 13.

Future Work

In FY 87, we plan to focus on improvements in all three research areas to enhance the applicability of the results to large structural systems. We will investigate a special time integration scheme to be implemented in GEMINI to solve the closed-loop dynamic equation and improved closed-loop model reduction techniques for the simulation procedure. We plan to conduct a systematic evaluation of a number of major sources of modeling errors in order to improve structural modeling techniques based on control stability considerations. We will also establish the component-mode synthesis modeling capability applicable to complex structural systems. With regard to the optimal decentralized control algorithm, we will continue our efforts toward efficient

computational algorithms and the enhancement of stability robustness.

Outside Contacts

We have had close interaction with the Strategic Defense Initiative Office (SDIO)-funded Controls Technology Research Program and with a closely related engineering research project in the Electronics Engineering Department at LLNL. We also had contact with R Program and the Precision Engineering Program at the Laboratory. We have constantly received guidance and review from the Dynamic Systems and Control joint thrust area leaders and other ME and EE professionals. We have received technical consultations from and conducted technical exchanges with the Aerospace Corporation, Dr. J. W. Young of SDRC, Inc.; Prof. R. K. Yedavalli of the University of Toledo, OH; Prof. Robert Taylor of the University of California, Berkeley; and Prof. J. R. Hutchinson of the University of California, Davis.

Financial Status

The total expenditure on this project for FY 86 was about $265 000.

1. S. C. Lu and S. E. Bumpus, "Research Issues in Structural Engineering for Control of Large Space Structures," *Engineering Research Annual Report FY85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985).
2. R. C. Murray, *GEMINI—A Computer Program for Two and Three Dimensional Linear Static and Seismic Structural Analysis*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-20338 (1984).
3. *CTRL-C User's Guide* (Systems Control Technology, Palo Alto, CA, March 1984).
4. B. Nour-Omid and R. W. Clough, "Dynamic Analysis of Structures Using Lanczos Coordinates," *Earthquake Engineering and Structural Dynamics* **12**, 565–577 (1984).
5. A. E. Yanchevsky and V. J. Hirvonen, "Optimization of Feedback Systems with Constrained Information Flow," *Int. J. Systems Sci.* **12**(12), 1459–1468 (1981).
6. O. Zienkiewicz, *The FEM in Engineering Science* (McGraw-Hill, New York, NY 1971).
7. J. W. Young, "A Structural Dynamics Approach to the Simulation of Spacecraft Control/Structure Interaction," Identification and Control Workshop of the 1984 American Control Conf., San Diego, CA, June 4–6, 1984.
8. D. T. Gavel and H. H. Woo, *Control of a Flexible Beam*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-20339 (1985).
9. W. C. Hurty, "Vibration of Structural Systems by Component Mode Synthesis," American Society of Mechanical Engineers, *J. Engr. Mechanics Division* **86**, 51–69 (1960).
10. J. W. Pastrnak, "Determination of Complex Criteria for Design, Structural Analysis, and Controls," *Engineering Research Annual Report FY 85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985).
11. S. E. Bumpus and S. C. Lu, *Demonstration of Component Mode Synthesis Using Lanczos Coordinates*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-20885 (in progress).
12. A Flashner, "An Orthogonal Decomposition Approach to Modal Synthesis," *Int. J. Num. Meth. Engr.* **23**, 471–493 (1986).
13. S. H. Wang, S. C. Lu, and I. K. Fong, *Decentralized Optimal Control of Large Flexible Structures*, presented at 29th Midwest Symp. on Circuits and Systems, August 11–12, 1986; Lawrence Livermore National Laboratory, Livermore, CA, UCID-94993 (1986).

Study of Complex Modes

J. W. Pastrnak
Nuclear Test Engineering Division

This eighteen-month study has been successful in providing the designer and analyst with qualitative guidelines on the occurrence of complex modes in the dynamics of linear structures, and also in developing computer codes for determining quantitatively which vibration modes are complex and to what degree. In this second year of study, the work has focused on predicting complex mode behavior. I have experimentally verified the presence of complex modes in a test structure, performed a finite element analysis of the structure with non-proportional damping, and formulated a partial differential equation to eliminate possible modeling errors.

Introduction

The study of complex modes (more formally: the study of complex vibrational modes of structures) is concerned with quantifying the way in which mechanical systems tend to oscillate in their natural modes of vibration.

As noted in the FY 85 *Engineering Research Annual Report*,[1] the key to describing the vibrational nature of a mechanical system (the eigenvectors) lies in the nature and distribution of the system's damping. For example, for an undamped system (i.e., one that does not have any energy dissipation mechanisms) the natural modes of vibration are called "real normal" modes. Theoretically, when a system is vibrating in a normal mode, the system's total energy is constant and the system will continue *ad infinitum* to pass through states of pure kinetic energy to pure potential energy at a rate equal to the frequency of vibration (eigenvalue).

In reality, though, all mechanical systems exhibit some form of energy dissipation in the form of damping. This damping produces eigenvectors with real and imaginary parts and the modes of vibration are appropriately called complex modes. For every complex eigenvector there will also exist a complex conjugate eigenvector, so that the projection of each counter-rotating eigenvector (phasors) leads to a real mode shape when superimposed. When the phase angle between the components of each eigenvector equals $N\pi$, where $N = 0, 1, 2...$, the mode shapes are identical to the undamped mode shapes. This special case results when the system is proportionally damped or more generally, as pointed out last year, when the system damping is in Caughey form.[2] Thus a real normal vibrational mode can be considered as just a special case of the general complex mode.

Technical Status

The concept of modal analysis has proven to be a very useful and efficient tool for analyzing the dynamics of linear structures.

This year's work has focused on predicting complex mode behavior, using three different approaches:

- Experimental verification of the presence of complex modes for a test structure.
- Finite element analysis of the above test structure with non-proportional damping.
- Formulation of a partial differential equation for a simple test structure to eliminate possible modeling errors introduced by discretization.

Experimental Verification

An experimental modal analysis was performed on a redesigned, damped cantilever beam. This year, however, the damper was chosen and then the beam was sized so as to make the damping forces of the same magnitude as the restoring forces due to the beam

stiffness. The damper used to approximate linear viscous damping was a voice coil from an industrial disk drive. By using the voice coil essentially backwards (that is, to drive the coil with a force and velocity as input and to dissipate the generated voltage and current as output) a linear viscous damper could be implemented. The idea of using a voice coil as a damper arose out of discussions with last year's consultant from U.C. Davis, Professor D. Margolis.[1]

The constitutive relations to describe a linear voice coil are:

$$e = rv \quad , \tag{1}$$

and

$$ir = F \quad , \tag{2}$$

where

r = voice coil (gyrator) modulus,
e = generated output voltage,
i = generated output current,
v = input velocity,
F = input force.

To use the damper, I first had to calibrate the voice coil to determine its effective damping coefficient C. The effective damping coefficient C is given by:

$$C = \frac{F}{v} = \frac{r^2}{R_v} \quad , \tag{3}$$

where

R_v = external resistance used to dissipate energy.

By measuring the voltage drop e, across R_v, and the driving velocity v, it was possible to calculate r using Eq. (1).

A second method of calibration, which led to more accurate results, came from measuring F/v directly. Figure 1 shows the laboratory set-up. A precision velocity transducer was used to measure the coil velocity while a standard load cell was used to measure the input force. Up to approximately 800 Hz, the value of C for displacements under 0.25 in. was found to be 0.015 lb-sec/in.

The next step was to demonstrate the effect of the damper experimentally. Using Bernoulli-Euler theory for beam bending, I calculated the location of the nodal lines* for the cantilever beam in its third bending mode. The damper coil was then attached to the beam at the nodal line closest to the free end of the beam (Fig. 2). The beam itself was constructed from a piece of homogeneous aluminum 66.22-in. long by 1-in. wide and 0.25-in. thick. At 16 measurement locations, evenly distributed along the length of the beam, 5.5-gram masses were added to simulate the mass of the accelerometer, except at the location where the damper was

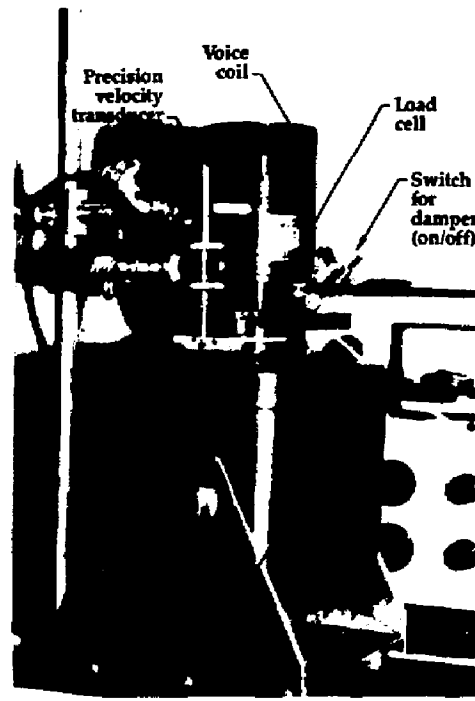


Figure 1. *Voice coil damper being calibrated.*

* Nodal lines are locations on the beam that at specific frequencies have no motion. (They are not to be confused with node points, which are points on the structure chosen for reference and usually, in this context, for measurements.)

mounted. The mass of the moving part of the damper was 27 grams.

The reason for attaching the coil at this location was to show that all the modes except the third bending mode would be affected by the damper. Figure 3 shows the cantilever beam with the damper attached and the data acquisition equipment used. As predicted, careful placement of the damper caused the first, second, and fourth bending modes to be affected, leaving the third mode unchanged (Table 1).

Examination of the animated mode shapes with PC-MODAL[3] on an IBM personal computer clearly showed the presence of the damper. The first bending mode was most affected, evidenced at the damper location, which lagged behind the other nodes. Phasor plots of the mode shapes showed virtually no difference for the third mode with the damper on or off. Figure 4 is a typical frequency response function from the IBM-PC and depicts the textbook quality of the actual data. The frequency response function is a complex quantity, in this case expressed by amplitude in the top plot and phase in the bottom plot. The locations of the peaks on the amplitude plot correspond roughly to the resonant frequencies of the beam. The actual resonant frequencies listed in Table 1 were found by curve-fitting the measured frequency response functions with a multi-degree-of-freedom complex exponential curve-fitter in PC-MODAL.

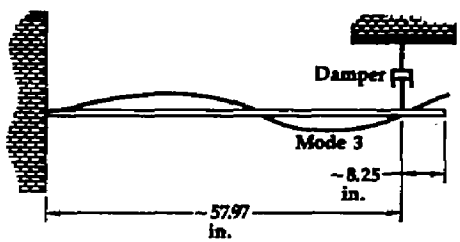


Figure 2. Third bending mode of a cantilever beam.

Finite Element Analysis

Using the same cantilever beam design (including the point masses at the nodes), I created a 16-element model using beam elements in GEMINI.[4] With this model I was then able to do a normal mode analysis using

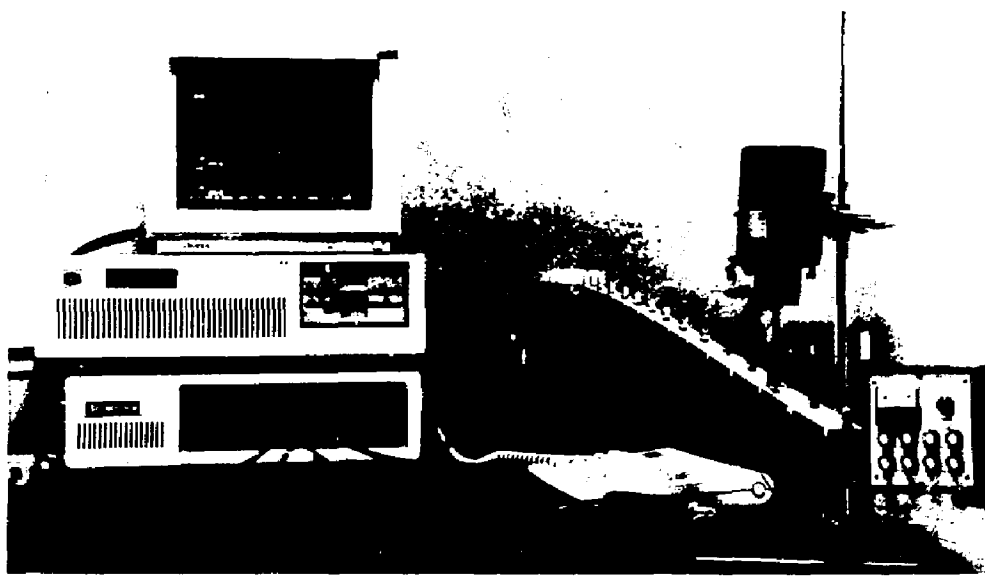


Figure 3. Experimental setup for the damped cantilever beam.

GEMINI on the Cray-1 computers. Then, using the program KREAMX (written for this project by Stan Bumpus), I was able to extract the undamped beam's mass and stiffness matrices from GEMINI. I then inserted into an empty damping matrix the dashpot coefficient from the voice coil calibration at the correct degree of freedom corresponding to the damper location. Using the CPLXMI program (also written by Stan Bumpus for this project), I then performed a complex eigenanalysis using these matrices. Deflections due to internal shear forces were included, but the rotatory inertia effects were not included because GEMINI overcompensates these effects. Table 2 gives the results of the GEMINI calculations as well as those from the CPLXMI program. These results show the frequencies obtained in both these analyses correspond well with those obtained experimentally (Table 1).

Partial Differential Equation Formulation

The third approach pursued was to use the partial differential equation (PDE) to solve the complex eigenproblem. The PDE approach was chosen to preclude any discretization errors that may be inherent in the finite element approach. Clearly, the most straightforward solutions for the PDE arise from the assumption of separation of variables:

$$w(x,t) = Z(x)W(t) \quad , \tag{4}$$

where $Z(x)$ and $W(t)$ are spatial and temporal functions, respectively.

The question arises: can a complex mode be separated into temporal and spatial parts exactly, when the damping may be a function of the spatial variable x and the temporal variable t?

Mode	Damper off		Damper on	
	Frequency (Hz)	Damping (% cr)	Frequency (Hz)	Damping (% cr)
1	1.67	4.35	1.52	23.3
2	10.63	0.61	10.60	1.38
3	30.20	0.65	30.20	0.61
4	58.66	0.51	58.60	0.57

Table 1. Comparison of first four measured bending modes, with and without the damper turned on.

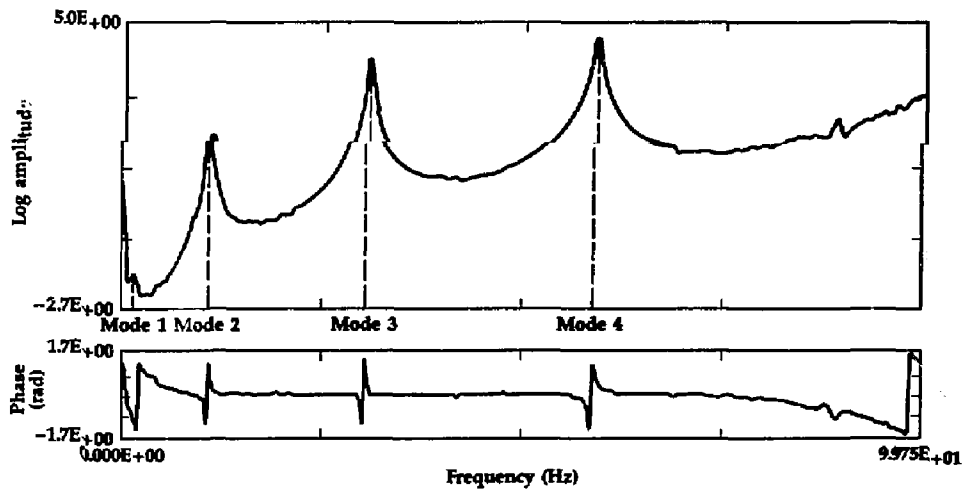


Figure 4. A typical measured frequency response function for the beam.

To help answer this question, I asked Professor W. C. Hurty from U.C. Los Angeles to look at the PDEs of a Bernoulli-Euler beam with a point damper attached. Professor Hurty's work has been documented in the form of a UCID report.[5] Dr. Hurty worked with the Bernoulli-Euler formulation of a beam in transverse bending:

$$m_l \frac{\partial^2 w}{\partial t^2} + C\frac{\partial w}{\partial t} + EI\frac{\partial^4 w}{\partial x^4} = 0 \quad , \quad (5)$$

where

$w(x,t)$ = transverse displacement,
m_l = mass per unit length,
EI = flexural stiffness,
C = coefficient of viscous damping.

He reported that when the damping term C in this formulation is not constant, and in fact may be a function of x, then separation of variables in the normal sense may not be possible. Instead, Professor Hurty offered an approximate solution by expanding the displacement $w(x,t)$ with a set of orthogonal functions. The details of this methodology and a sample application can be found in Ref. 5.

A second plan to aid in understanding complex mode behavior was to have been in modeling the free vibrations of a mechanical system using traveling transverse, longitudinal, and Rayleigh[6] waves. With this type of continuous formulation, a resonance can be accounted for by the *constructive* interference of two opposite traveling waves. Likewise, an anti-resonance explained by the *destructive* interference of two opposite traveling waves. As the speed of the wave varies due to non-constant dispersion in the material, it is conceivable that as the waves superpose, a smearing or delay may occur and thus give rise to an observed complex mode! Unfortunately, the scope of work for this program did not allow for detailed traveling wave analysis.

Conclusions

From the results of FY 85–86 work, it is concluded that if the damping in a structure is not uniform, the resulting vibrational mode shapes may be complex. Causes of non-uniform damping include point-located dampers (dashpots), non-uniform viscous vibration suppression coatings, rate feedback actuators, and friction in joints. For any linear system containing any of these elements, the designer is strongly cautioned to be prepared to analyze the system with a complex eigenanalysis if accuracy is important.

The original goal was to provide the structural designer and analyst with guidelines on when to expect complex modes. The conclusions above will serve this purpose; however, these statements are only qualitative criteria at best and do not give a quantitative measure of the degree of complexity of the resulting complex modes. For this reason, we developed the computer codes CPLXKI, CPLXMI, and KREAMX to determine more concisely which modes were complex and to what degree. Our work has provided these tools to the Laboratory as well as to the outside engineering community.

A second goal of this work was to make engineers and scientists more aware of the idea of complex modes. Professor Taylor at U.C. Berkeley now has a graduate student working on a complex eigensolution package for their previously normal mode finite element work. The Modal Analysis Lab of the Dynamic Test and Analysis Group in Nu-

Table 2. Comparison of calculated modes, undamped and damped.

	Damper off		Damper on	
Mode	GEMINI frequency (Hz)	CPLXMI frequency (Hz)	CPLXMI frequency (Hz)	CPLXMI damping (% cr)
1	1.65	1.67	1.62	25.5
2	10.60	10.61	10.60	0.80
3	29.80	29.81	29.80	0.0001
4	57.89	57.89	57.89	0.106

clear Test Engineering Division performs all modal test analysis using complex modes. Another project in the Engineering Research Program at LLNL is trying to use Lanczos vectors to perform a complex eigenanalysis.

Future Work

Though this work will not continue into FY 87, this report will serve as a benchmark for the work to date, and also as a guide for further efforts in the near future. I expect that the KREAMX and CPLXMI programs will be used by various LLNL personnel when it becomes necessary to model mechanical systems that contain point-located viscous dampers. The answers to questions such as "How is a dashpot coefficient related to the degree of complexity for any given mode?" or "How much phase variation can be expected between structural nodes for a given mode before critical damping is reached?" are unknown and are left for the future.

Acknowledgments

I would like to acknowledge the following individuals for their support that contributed to the success of this project for FY 86. First, I would like to thank Howard Woo from the Engineering Research Program for his guidance and funding. Thanks also to Joe Weaver for the use of the Modal Lab facilities, Gerald Goudreau from the Methods Development Group, and special thanks to Stan Bumpus for writing the KREAMX program as well as his patience in helping me with GEMINI. Thanks also to the management of the Nuclear Test Engineering Division for allowing me time off to work on this project.

Financial Status

The total expenditure on this project for FY 86 was $134 000.

1. J. W. Pastrnak, "Determination of Complex Mode Criteria for Design, Structural Analysis, and Controls," *Engineering Research Annual Report FY 85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985), pp. 32–33.
2. T. K. Caughey, "Classical Normal Modes in Damped Linear Dynamic Systems," *J. Appl. Mech.* 27(2), pp. 269–271 (1960).
3. *PC-MODAL Users Manual* (Modal Test Associates, San Jose, CA, 1985).
4. S. E. Bumpus and R. C. Murray, *GEMINI—A Computer Program for Two and Three Dimensional Linear Statics, and Seismic Structural Analysis—Cray Version*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-20338 (1986).
5. W. C. Hurty and J. W. Pastrnak, *A Brief Study of Complex Viscous Damping*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-20913 (1986).
6. R. Wasley, *Stress Wave Propagation in Solids* (Dekker, New York, NY, 1973), pp. 55–67.

Chip Science: Basic Study of the Single-Point Cutting Process

R. R. Donaldson
Precision Engineering Program

C. K. Syn and J. S. Taylor
Materials Fabrication Division

R. A. Riddle
Engineering Sciences Division

During FY 86 we measured the wear that diamond tools sustained during the cutting of electroless nickel (eNi). We were able to detect wear at previously unattained levels, down to 100 Å, and we found that the tool wear resulted in a burnishing action after a relatively short cutting distance. To provide a more direct connection between computer-based modeling and experimental measurements, we also performed macroscopic cutting tests on a well-characterized aluminum material. The results showed good agreement between calculated and measured cutting forces. We also completed the installation of a new high-resolution cutting force dynamometer.

Introduction

Our long-term objective is to develop a fundamental understanding of the single-point cutting process, with an emphasis on attaining a high degree of accuracy and a smooth surface finish. Our pursuit of this objective draws on three areas of technical strength at LLNL:

- Precision engineering, for accurate cutting experiments and measurements of geometry and surface roughness.
- Materials science, for insight into the deformation process, aided by a wide variety of characterization techniques.
- Computer-based numerical analysis, for the ability to check our basic understanding by comparing calculations and experiments, leading ultimately to a predictive capability.

During FY 86, we conducted an extensive diamond tool wear test on the cutting of electroless nickel, using a preparation, previously tested in FY 85, that had yielded the lowest surface roughness. We also performed orthogonal cutting experiments on a well-characterized 7075-T6 aluminum material, under experimental conditions designed to provide direct comparisons with our finite-element modeling work. Additions were also made to our experimental test and characterization facilities.

Technical Status

Electroless Nickel Cutting Tests and Characterization Studies

During FY 85, we made an extensive study[1,2] of electroless nickel (eNi), during which we varied the percentage of phosphorous and the heat treatment conditions. The study measured the effects of these parameters on the eNi microstructure, hardness, diamond-machined surface roughness, and tool wear. The eNi preparation yielding the lowest surface roughness (13 wt% P with a 2-hr, 200°C heat treatment) showed no detectable tool wear (as evidenced by SEM and optical microscopy) for the 2-mile cutting distance used in the test. In the present study, we extended the cutting distance for the optimum eNi preparation until definite wear was observed, with additional measurements to quantify the amount of wear. We also used two tools, chosen to have differences in the fracture strength and fracture toughness of their diamonds. A complete report of this test has been prepared.[3]

Cutting Test. Of the two diamond tools that were selected (from the six used in the previous study), one had the least, the other the greatest concentration of nitrogen-

platelet impurities, as determined from their infrared (IR) absorption characteristics. The tools were round-nosed, with radii of 0.030 in. After careful resharpening, these tools were used in the Precision Engineering Research Lathe (PERL)[4] to cut sets of 2-in.-diam copper disks that had all been plated with eNi in a single batch. The test sequence consisted of multiple passes of cutting and characterizing, beginning with examination of the tools before use. A cutting pass for a given tool consisted of facing each of four disks once, after which the tool was removed and characterized for wear, and the surface roughness of the four machined disk surfaces was measured. Each facing cut was 100 μin. deep, with a spindle speed of 1000 rpm. A light oil was used as a cutting fluid. The feed rate was 100 μin. per revolution, which yielded a cutting distance of about 2500 ft per part (10 000 ft per pass). At each installation, care was taken to align the tool shank parallel to the Z slide travel within 0.1 mrad, to assure that the same region of the tool tip was exposed to wear.

Characterization Techniques. The surface roughness was measured with a Talystep profiler, using a hachet-shaped diamond stylus that had a 4-μin. tip radius in the scanning direction. This instrument was equipped with an LLNL-designed data acquisition computer system that provided data storage, optional detrending, and calculation of rms amplitudes. Short scan lengths (500 μin.) were selected from longer measurements (4 mil) to suppress longer-wavelength contributions from machine and environmental sources. This emphasized changes, due to tool edge wear, in the surface profile across the individual feedmarks. Typically, measurements were made at five equally spaced radial positions on each disk, although a larger number of measurements were made over the first 1000 ft of cutting to examine the initial stages of tool wear.

The tools were examined with an optical microscope at magnifications up to 500X, in normal and Nomarski modes, and with an SEM at magnifications of 2500 to 10 000X. The tools were sputter-coated with a 50- to 100-Å layer of gold-palladium to avoid charging effects in the SEM; the layer was removed before the next cutting pass.

Another method used to track the tool wear was a tool edge replication technique, during which the tool nose made a plunge cut into an oxygen-free, high conductivity (OFHC) copper sample. This technique was based on an observation, made during the FY 85 study, that the "signature" of the tool edge across a feed-groove was repeated over nearby grooves within a few nanometers (1 nm = 10 Å = 0.04 μin.). A plunge cut replica was made initially and after each machining pass. The depth of the plunge cut was twice the eNi cut depth, so as to include adjacent unused edge regions for reference. A variation on this technique was added later in the test series: the tool was fed laterally for about ten spindle revolutions, with a relatively coarse feed per revolution before retraction. This demonstrated the reproducibility of the tool edge signature and added an angular (tilt) orientation to the tool-tip wear pattern. The grooves resulting from the plunge cuts were examined with an optical microscope and measured by a Talystep profiler.

Experimental Results. Figure 1 shows a rapid rise in the surface roughness, from about 5 to 30 Å rms, that occurs within the first 1000 ft of cutting distance for both tools. Figure 2 shows a more gradual rise to 90 Å rms, with significant local fluctuations, over the duration of the test (five passes for the tool designated as L-10 and seven passes for L-9).

Figure 3 is a sampling of the eNi surface profiles cut with tool L-9 at various distances; the figure aids in the interpretation of Figs. 1 and 2. Each plot is about five feedmarks wide. A comparison of profiles (a) and (b) clearly shows a significant change in the tool edge signature in less than 2000 ft of cutting distance. Profile (c) has a smaller rms amplitude than (b), and a more jagged shape, while in (d), only 80 ft later, the profile is considerably smoother. This sequence suggests that a microfracture of the tool edge occurred just prior to profile (c), in a location that reduced the rms amplitude, and that a

chemical reaction and/or dissolution mechanism is also operative that tends to smooth the fractured edge. A comparison of profiles (b) and (e) shows that the amplitude of the latter is only about 50% larger, despite being at 20 times the cutting distance. Finally, (f) is indicative of the variety of periodic tool edge signatures that were observed.

Figure 4 shows a montage of optical micrographs of the bottom of the plunge-cut replica grooves made between the eNi machining passes; the cutting distances are labeled in miles. The figure shows that the tool tip is producing a set of closely spaced parallel lines or grooves that increase in number with the cutting distance. The spacing between the lines (100 μin.) is the same as the feed per revolution (feedmark width) used in cutting the eNi samples. The location of the lines is at or beyond the trailing edge of the cut, where no tool-to-work contact would be expected unless wear had changed the shape of the tool tip.

Figure 5 shows the Talystep profile of the lateral-feed replica groove for each tool after the final cutting pass. The tool motion in Fig. 5 is from right to left. It is evident that the circular tool tips have receded by about 750–1000 Å (3–4 μin.) to a flattened shape. Also, the flattened region exhibits a highly periodic structure consisting of 5–7 grooves, each 100 μin. wide, equal to the feedmark spacing. Thus it is evident that wear of the tool edge has resulted in tool-to-work contact over some half-dozen feedmarks, with the final interaction being one of burnishing rather than cutting. The flattened region exhibits a small but definite slope that is probably a consequence of the workpiece and tool deforming in response to the thrust component of the cutting force (measured at approximately 10 g), although no calculations have been performed to quantify this effect.

Figure 6 shows a series of Talystep profiles taken of the replica grooves of Fig. 4. The figure depicts the evolution of the trailing edge wear notches as a function of cutting distance. Figure 7 shows a plot of the edge recession distance vs cutting distance. Edge recession causes figure error in a machined optic (e.g., if the tool were used for a single long facing cut on a large eNi optical flat).

A final confirmation of the presence of trailing edge grooves is the five feed-spaced marks (arrows) along the cutting edge of tool L-10 (Fig. 8, a composite of several SEM micrographs).

The formation of regular feed-spaced wear notches has been observed on other tool materials, such as high-speed steel and cemented carbides, at much larger magnitudes,[5] and has also been observed when diamond tools have sustained severe wear

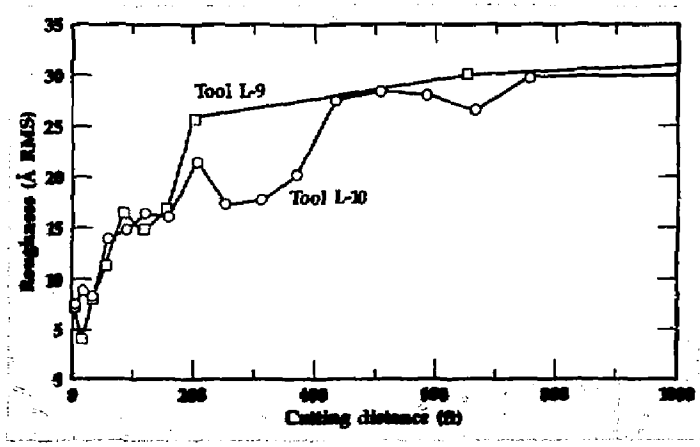


Figure 1. Increase of the rms machined surface roughness over the first 1000 ft of cutting distance, for both cutting tools.

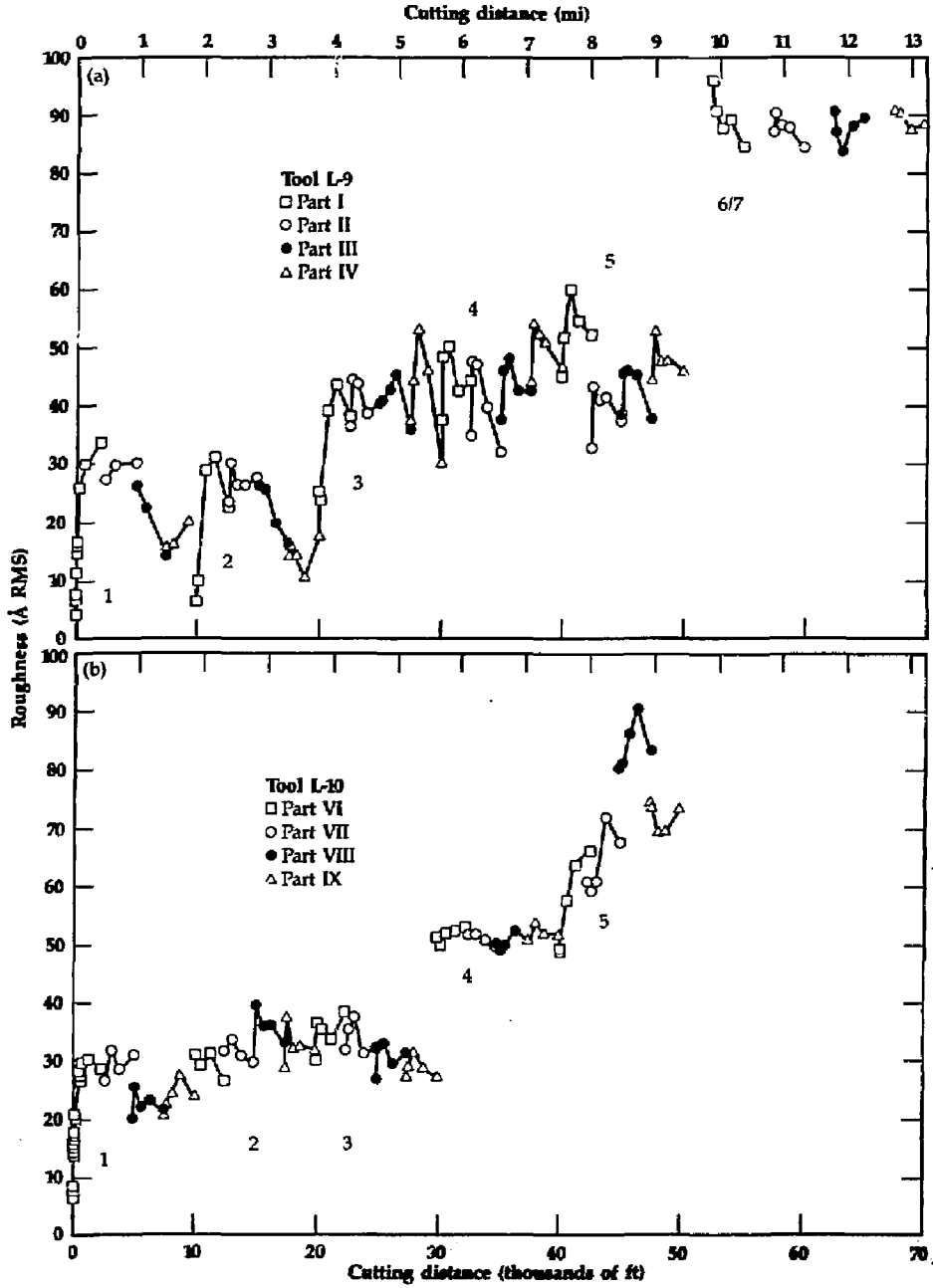


Figure 2. Variation of the rms surface roughness vs cutting distance for the full duration of the wear study, for (a) tool L-9 and (b) tool L-10. Each pass, consisting of cutting four parts in succession, is labeled with a pass number (1 to 7 for tool L-9 and 1 to 5 for tool L-10).

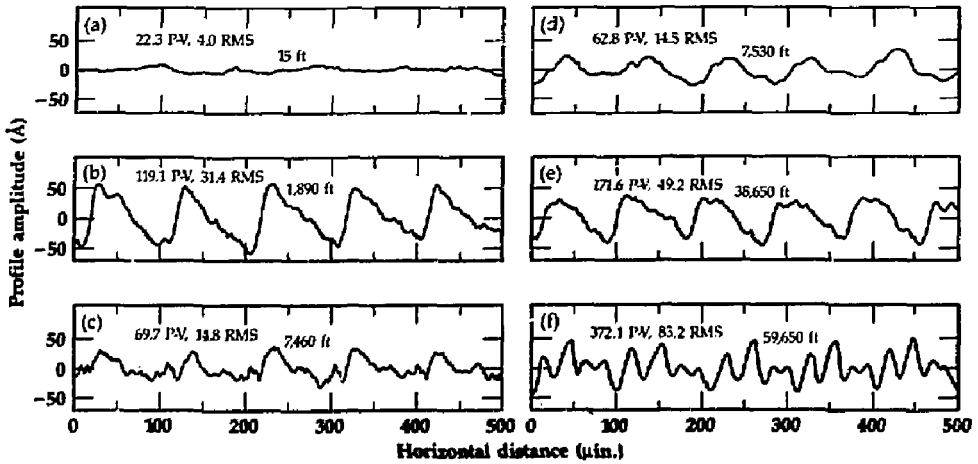


Figure 3. Talystep profiles of the feedmarks cut into the eNi surface by tool L-9 at the cumulative cutting distance shown. Each profile is about five feedmarks wide and the tool advance is from right to left. The peak-to-valley and rms amplitudes are in angstroms. Note the change of vertical scale for (e) and (f).

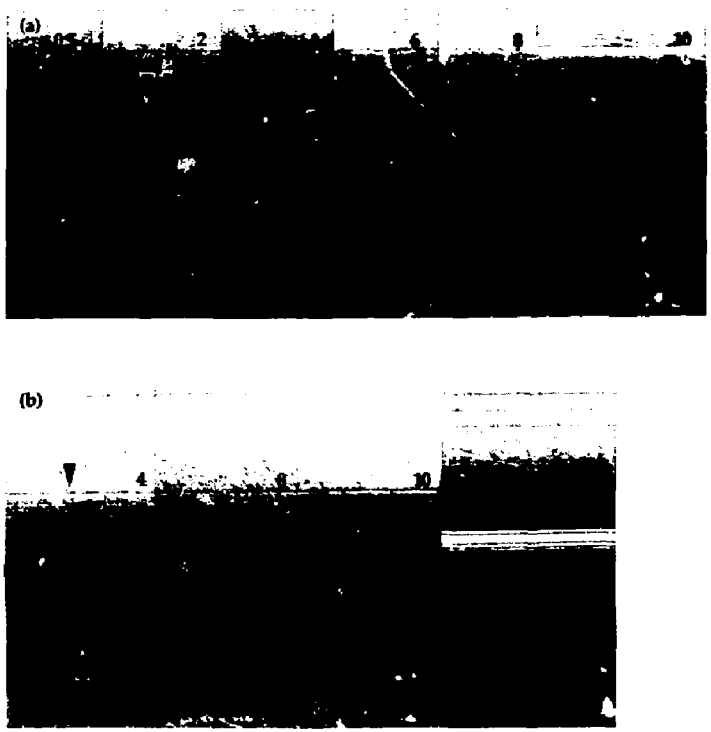
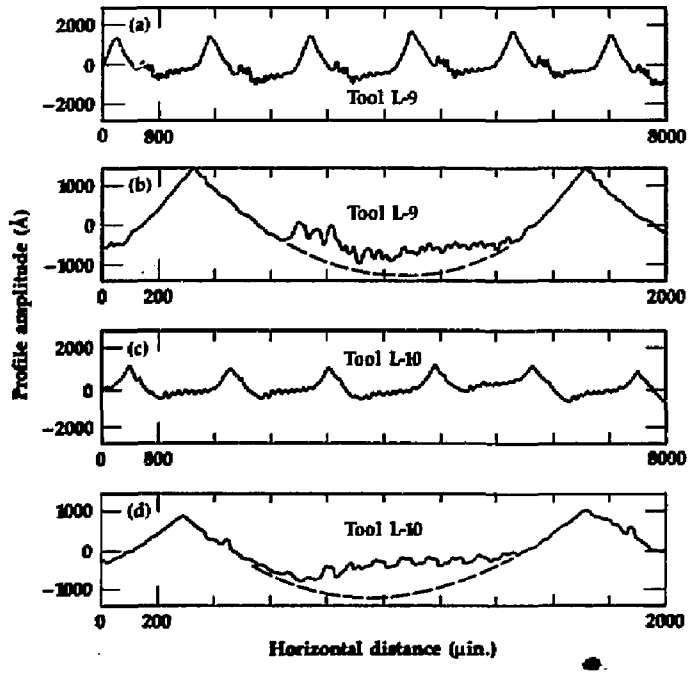


Figure 4. Montage of photographs (taken with the optical microscope) of the trailing edge side of the replication grooves for (a) tool L-10 and (b) tool L-9. The number in each photograph represents the cutting distance in miles; clearly visible is the increase, with cutting distance, of the size and number of lines caused by trailing edge tool notches. The photographs were aligned by using the lines (indicated by the arrows) caused by a nick in the tool. The leading edge of the tool is upward in the photographs. The scaled bar length is given in micrometers.

Figure 5. Talystep profiles of samples that have been laterally fed to replicate tool edges. The profiles show the full set of feedmarks at lower magnification and one feedmark at higher magnification, for tool L-9 after pass 7 (a, b), and tool L-10 after pass 5 (c, d). Tool motion is from right to left at a feed rate of about 1.5 mils/revolution. The dashed curve is an estimate of the original tool contour.



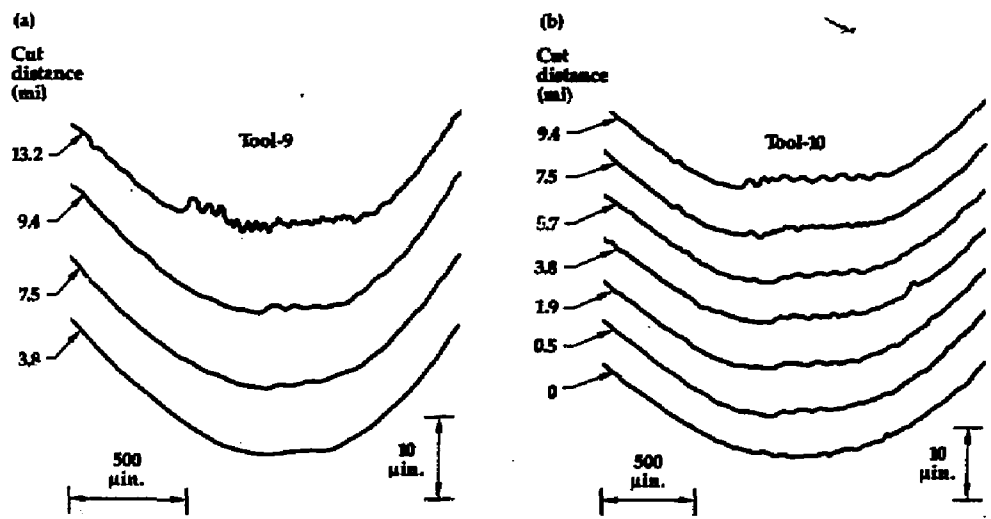


Figure 6. Talystep profiles of the replication grooves shown in Fig. 4. The profiles show the progression of the tool tip recession vs the cutting distance for (a) tool L-9 and (b) tool L-10. The uppermost profiles correspond to those in Fig. 5.

during the cutting of mild steel.[6] To our knowledge, it has not been reported previously for diamond tools at the very small scale observed in the present study.

Based on the IR absorption data, tool L-9 would be expected to have higher hardness but lower fracture toughness than tool L-10. In Figs. 4 and 6, tool L-9 develops wear notches more slowly than tool L-10, with a lower initial wear rate slope in Fig. 7. However, in Figs. 2 and 7, tool L-9 also shows more abrupt increases in roughness and edge recession amplitude, which may be due to fracture of the edge. Thus the general trend is in the expected direction, although the statistical data base is obviously very small.

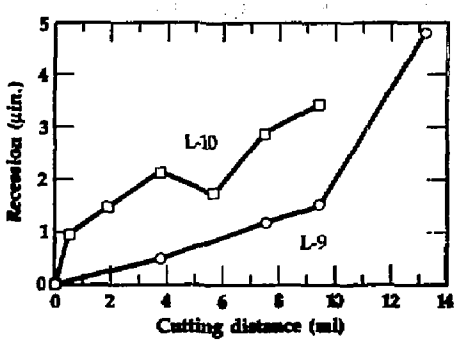


Figure 7. Plot of the tool edge recession distance vs cutting distance for both cutting tools, based on the Talystep profiles of Fig. 6.

Two findings of the extended eNi wear study were unexpected, and are deserving of further investigation. First, for the nominally identical eNi preparation, the tool wear was greater than was projected from the FY 85 study. As discussed further in Ref. 3, this may be explained by a tendency of eNi baths to fluctuate, depositing layers containing higher or lower phosphorous concentrations. (The wear rate of diamond is known to increase when the P content drops below about 11 wt%.) If confirmed, improved techniques for fine control of the plating bath parameters might be possible.

The second unexpected result was the evidence indicating a rubbing or burnishing interaction at the trailing edge of the tool cutting zone. This is contrary to the widely held view that the mirror finish produced by diamond tools is a result of clean cutting with an extremely sharp edge, without burnishing. The clean-cutting model may still be valid at small cutting distances (and over relatively long distances for more benign work materials such as copper). However, Figs. 2 and 4 show that multiple wear notches occur when the surface roughness is in the 30- to 50-Å rms range (hence, "mirror-like"), which shows that burnishing and mirror finishes are not incompatible. This relates to a question of considerable theoretical and practical interest: what is the minimum attainable depth of cut for a given state of tool wear; i.e., at what uncut chip thickness will the work material flow under the tool edge with-

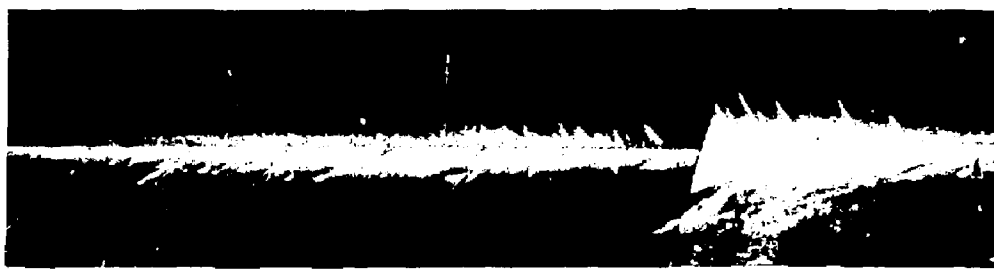


Figure 8. SEM photographs of the trailing edge region of tool L-10, showing the presence of wear notches (five arrows). Note that the notches have the same spacing as the feedmarks on the eNi surface, 100 μin. or 2.5 μm. The single right-hand arrow indicates the nick used for alignment in Fig. 4.

out being removed? Further study of this problem has begun and will be continued in FY 87.

Numerical Modeling of the Chip Formation Process

Introduction. We have numerically modeled the chip formation process using the finite element program NIKE2D. Our efforts have been focused on three areas:

1. The choice of the failure criterion in analyses of orthogonal metal cutting. The correct value of the effective plastic strain-to-failure, ϵ_{pf}, along the tie-break slideline has been derived using concepts from fracture mechanics.
2. The modeling of the orthogonal metal cutting experiments, particularly with regard to choosing the best work-hardening work-softening (WHWS) constitutive model for simulating the measured shear angle, cutting forces, and chip morphology.
3. Preliminary indentation studies aimed at (a) quantifying the effect of the tool nose radius and tool stiffness on the load required for machining and (b) determining the minimum possible depth of cut.

In each of these areas progress has been made in understanding how the mechanics of material deformation relate to the chip formation process.

The Failure Criterion in Orthogonal Metal Cutting. In the finite element cutting analyses, we modeled chip formation and crack growth with a "tie-break slideline." A tie-break slideline consists of two sets of nodal points, initially coincident, that are allowed to separate when a specified criterion is met. Values of strain in the elements adjacent to the tie-break slideline are extrapolated to obtain values at the double-node points. When the value of effective plastic strain at a nodal point reaches a critical input value, the "tie" between the double nodes on the slideline is "broken," and material separation occurs between the chip and the workpiece surface.

The theoretical correlation between the critical effective plastic strain-to-failure, ϵ_{pf}, along a tie-break slideline and the J Integral has been derived. The J Integral is a measure of the energy required for crack growth in an elastic-plastic material, and is the subject of a widely accepted standard for experimental measurement. The correlation between the J Integral and the ϵ_{pf} has also been demonstrated numerically in an orthogonal cutting analysis, using a 60° rake angle tool and aluminum 7075-T6 workpiece material.[7]

A 60° positive rake angle was chosen for this demonstration to limit the amount of plastic deformation in the chip as the tool progresses (Fig. 9). The large mesh distortions found in the analyses of smaller rake angles make the verification of the line integral calculations difficult. The material properties of the aluminum alloy used in the analysis are shown by the solid curve in Fig. 10. The critical strain value for the material is 0.15.

Figure 11(a) shows the J Integral values as a function of tool travel (i.e., cutting distance with a constant tool velocity). A node

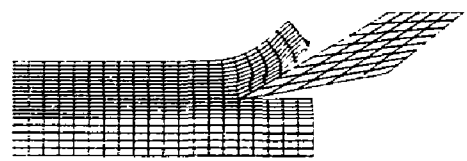


Figure 9. The NIKE2D chip deformation model, with a 60° positive rake angle, used to correlate the J Integral with the effective plastic strain-to-failure chip separation criterion.

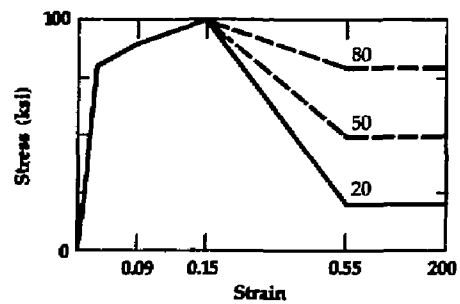


Figure 10. The work-hardening work-softening (WHWS) true stress-strain curves used for the NIKE2D calculations of chip flow in aluminum 7075-T6. The solid line was used for the calculation shown in Fig. 9.

opening occurred at 0.068 in.; the negative values preceding the opening and after 0.080 in. appear to be a code anomaly, in that the reversed sign occurs when there is an interpenetration between a tool tip node and a tie-break slideline node. If the sign change is ignored, the equivalent stress intensity factor K_I can be found from

$$K_I = \left(\frac{|J|E}{1 - \mu^2} \right)^{1/2} ,$$

where μ is Poisson's ratio and E is the elastic modulus of the material. The resulting values of K_I are shown in Fig. 11(b). The fact that, for a single value of ϵ_{pf}, K_I remains nearly constant as the chip formation progresses indicates a unique relationship between the two failure criteria.

These analyses of the failure criterion in metal cutting have shown that the effective plastic strain-to-failure along a tie-break slideline and the J Integral are correlated. Thus, if we know the J Integral value for a material, we can calculate the effective plastic strain-to-failure. Also, the exercise has shown that for most practical cutting analyses, where the rake angle is near zero, the mesh distortion is too large to perform the J Integral calculation accurately; hence, the effective plastic strain-to-failure is the best failure criterion for numerical implementation at the present time.

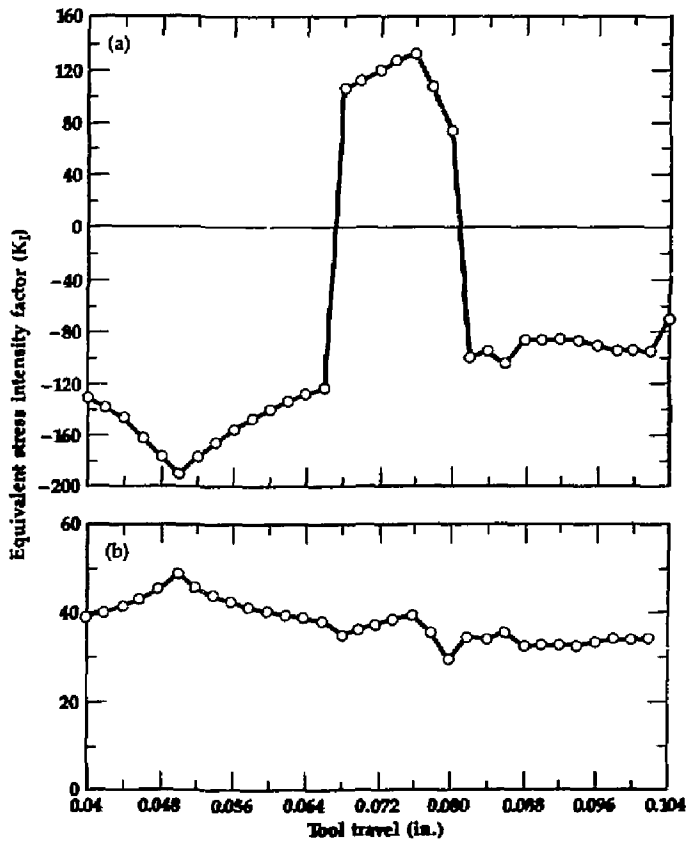


Figure 11. The variation of (a) and (b) vs the NIKE2D tool travel: (a) the J Integral value and (b) the calculated value of the stress intensity factor K_I corresponding to (a). A double node opening along the tie-break slideline occurs at 0.068 in.

Orthogonal Metal Cutting Experiments and Analyses. Orthogonal metal cutting experiments have been conducted on 0.125-in.-thick aluminum alloy blanks with high speed steel tools of various rake angles, using uncut chip thickness depths of 0.005, 0.010, and 0.020 in. The blanks were made from aluminum 7075-T6 compact shear specimens that had been tested previously for failure behavior under mixed mode loading.[8] While the experimental data reduction and the corresponding analyses are not yet complete, several important facts have emerged.

One conclusion is that the deformation in the chip and the chip shape are very dependent upon the stress-strain relationship beyond maximum load, particularly for the analyses of the near-zero-degree rake angle tool cuts. This idea is further clarified by considering several WHWS constitutive models for the aluminum material, as shown by the solid and dashed curves in Fig. 10. Here the elastic and plastic material properties are identical up to the point of maximum load; thereafter the stress is assumed to soften to constant load levels of 20, 50, and 80 ksi. This material behavior assumption for aluminum is borne out by actual experimental data for OFHC copper.[9]

In light of these assumptions, we examined the case of a $-10°$ tool rake angle (Fig. 12), where the shear angle formed by the tool tip and the material changes with the constant work-softening (WS) stress levels. For a WS level of 20 ksi, the shear angle in Fig. 12(a) is approximately 70°; for WS values of 50 and 80 ksi [Figs. 12(b) and 12(c)], the shear angles change to 45° and 30°, respectively. These results are preliminary because no attempt was made to include the effect of friction between the tool and chip (which tends to cause additional convergence problems with NIKE2D).

Photographs of chips forming during our orthogonal cutting experiments (Fig. 13) show the shear angle to be 45° or less, making the larger values of the WS stress appropriate. These photographs also allow us to estimate the depth of plastic deformation under the machined surface by examining the distortion of the milling lines on the transverse surface of the specimen blanks. The formation of the segmented chips of Fig. 13 results in large periodic oscillations in the cutting force. Figure 14 compares the peaks and valleys of the measured and computed cutting forces; measurement and computation are in good agreement.

An examination of the analytical and experimental results leads to the suggestion that if the WS stress levels and friction factor

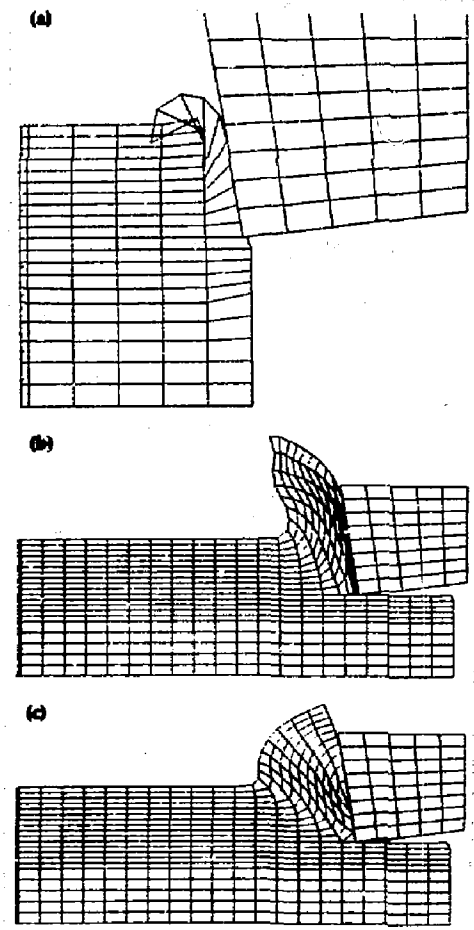


Figure 12. Three calculations of chip flow in 7075-T6 aluminum, using the true stress-strain curves of Fig. 10 with differing work-softening end-points: (a) 20 ksi, (b) 50 ksi, (c) 80 ksi. Result (b) is closest to the experimental results of Fig. 13.

are determined from the cut with the most negative rake angle (−10°) and the value for ϵ_{pf} is determined from the most positive rake angle cut (30°), then for all intermediate rake angle cuts the material properties should be adequately determined.

The orthogonal metal cuts provide an important opportunity to compare experimental and analytical results, and should provide valuable insights into the processes occuring in chip formation in three-dimensional cutting.

Indentation Studies. Analyses in which the side of a workpiece is indented by a rounded tool edge have been initiated, for a variety of reasons. First, attempts to analyze chips formed with a rounded tool edge have not been successful with NIKE2D, compared to chips formed with mathematically sharp tool edges, because of difficulties with the tie-break slideline position relative to the tool edge. The indentation studies do not employ a tie-break slideline. While this limits how far the tool can travel before the code is halted by excessive element distortion, it is possible to achieve significant plastic deformation of the work material, which can then be compared with an experiment. Second, by reducing the ratio of the uncut chip thickness to the tool edge radius, we can study the problem of the minimum depth of cut.

Figure 15 shows two indentation calculations, using an uncut chip thickness of 0.010 in. and tool edge radii of 0.001 and 0.002 in. The material model is the curve of Fig. 10 with a WS stress of 50 ksi. The friction factor between the tool and workpiece is 0.1. The deformation clearly proceeds to the onset of shear band formation before a convergence failure. Figure 16 shows the associated cutting forces, with the force required for the larger tool tip radius being about 10% higher. The analysis is presently ahead of the experiment, which has not yet been performed.

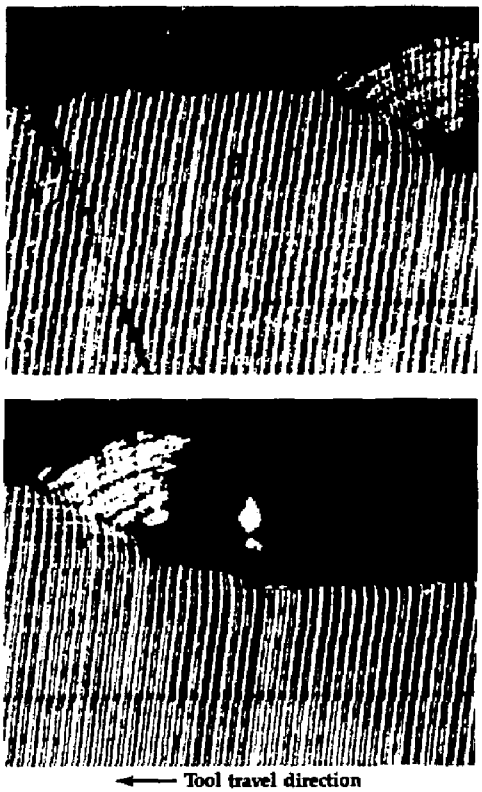


Figure 13. Photographs from the 7075-T6 aluminum orthogonal cutting experiments: (a) chip formed by a 0.020-in.-deep cut; (b) evidence of subsurface plastic flow provided by the milling cutter marks remaining from the preparation of the samples.

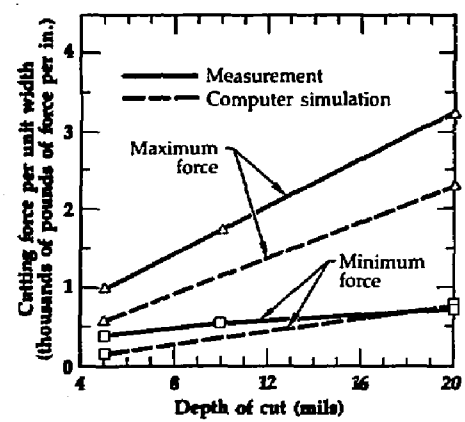


Figure 14. Maximum and minimum oscillatory force per unit chip width observed experimentally in the orthogonal cutting of aluminum 7075-T6, with comparisons to the NIKE2D chip flow analysis for the same material.

Figure 17 shows the results of two analyses of the minimum depth of cut. For a 0.001-in. tool edge radius, the 0.0001-in. depth of cut [Fig. 17(a)] results in a burnishing action without chip formation. Figure 17(b) shows the contours of effective stress for a 0.0004-in. depth of cut, where chip formation is in evidence. Thus, for this material model, the minimum depth of cut is less than 40% of the tool edge radius. Again,

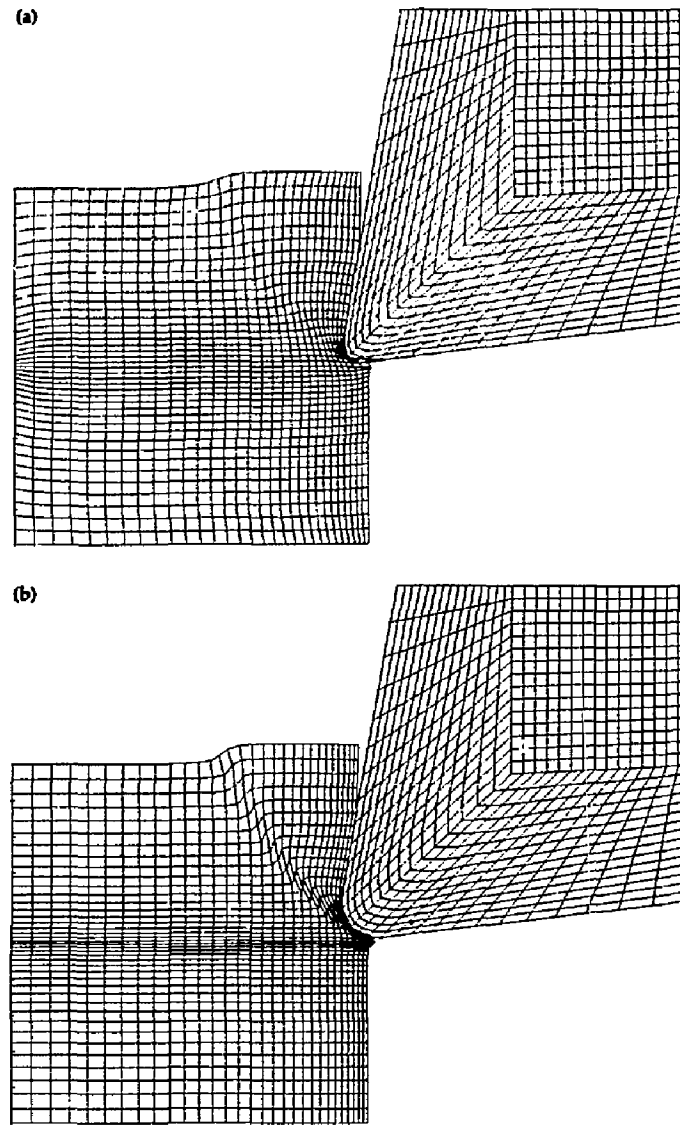


Figure 15. Analysis of an aluminum 7075-T6 work piece indented with a *round-edged 10° rake angle* tool, using NIKE2D without a tie-break slideline for chip separation, for a cut depth of 0.010 in. (a) 0.002-in. tool edge radius, (b) 0.001-in. tool edge radius.

the supporting experiments have not yet been performed.

NCSU Modeling Effort. The modeling work at North Carolina State University has focused on two areas: (1) improving the NIKE2D cutting model, and (2) studying an Eulerian approach to improve the Lagrangian NIKE2D model.

Previously,[1] the NCSU group had extended NIKE2D to include thermal effects caused by the plastic deformation of the elements. In designing the model, the group assumed adiabatic heating within each element. To model shear-banding deformations, materials that exhibit WHWS behavior and temperature dependence are required, a combination that was not available in NIKE2D. This feature was added to the "material type 4" option of NIKE2D, allowing general stress-strain curves to be input at eight different temperatures, with two-dimensional interpolation. The new routines were tested by use of a temperature-dependent, bi-linear, stress-strain example, which is also allowed under the existing "material type 3" option. Essentially identical results were demonstrated with both options. During testing, a difficulty with an existing subroutine input was discovered and corrected. Listings of the modified subroutines have been provided to LLNL.

In considering how to improve the modeling accuracy of temperature effects, the NCSU group examined the problems of including the internal conduction between the elements, and adding the frictional heating that is caused by the chip sliding along the rake and flank faces of the tool. The proper treatment of frictional heating requires identification of the nodes in contact with the tool surface and their relative velocities, neither of which is immediately available with NIKE2D. After a detailed examination, it was decided that the modifi-

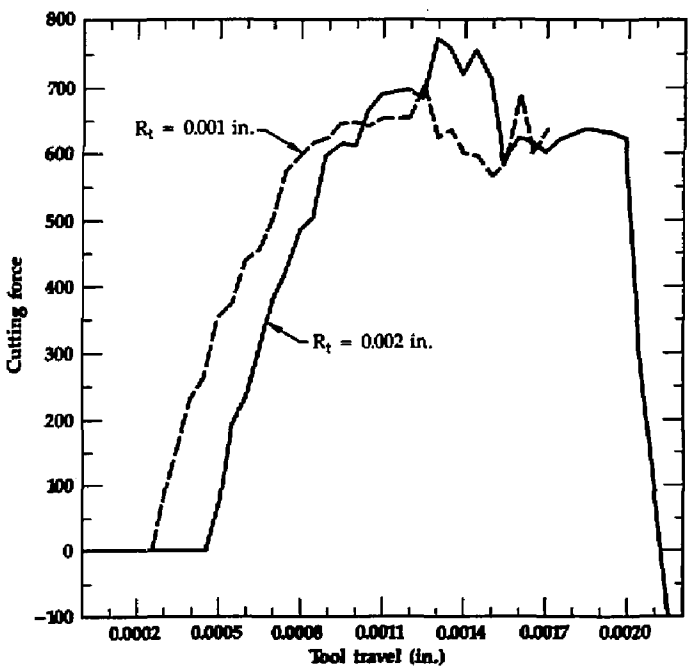


Figure 16. Variation of cutting force with tool travel distance for the indentation calculations shown in Fig. 15.

cations would require an effort exceeding the available resources. Instead, the group elected to combine their work to date on NIKE2D with an Eulerian code written as part of an NCSU doctoral thesis on chip flow.

The Eulerian code, which computes steady-state solutions for visco-plastic materials, has several advantages compared to a Lagrangian code such as NIKE2D. For example, even with round-edge tools, the chip separa-

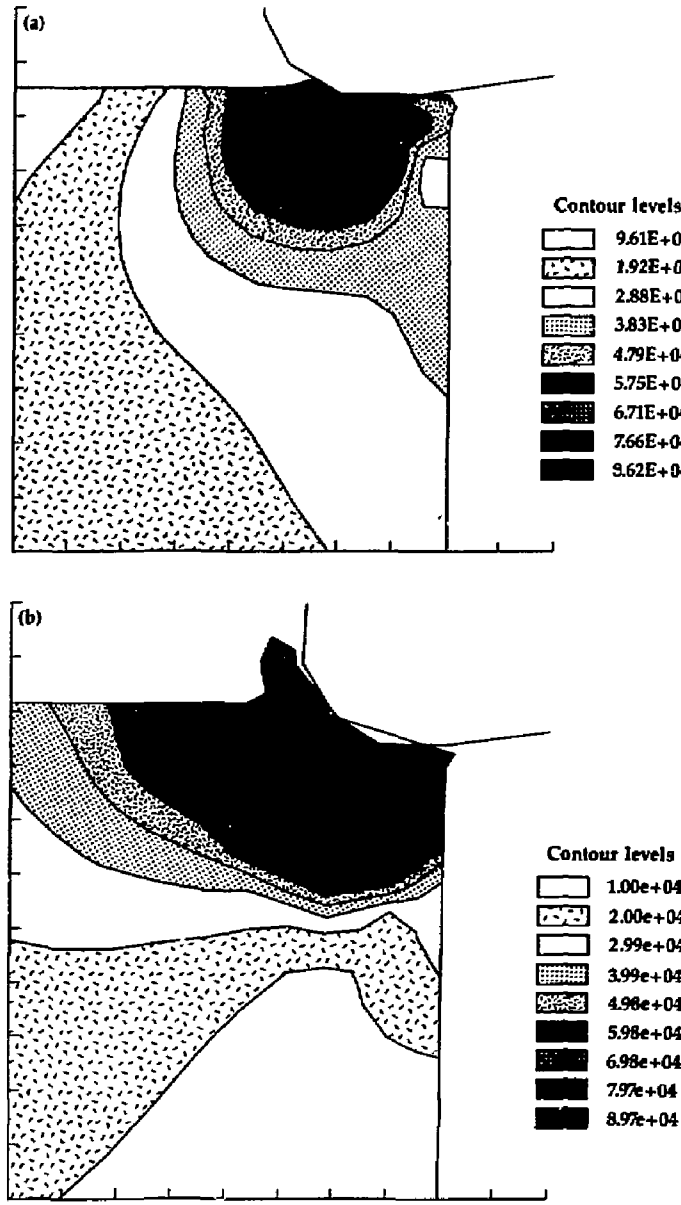


Figure 17. Contours of effective stress (in psi) during a minimum depth of cut simulation. The material is aluminum 7075-T6; the tool edge radius is 0.001 in.; the cut depths are (a) 0.0001 in., (b) 0.0004 in.

tion appears naturally as a stagnation point, avoiding the difficulties of a tie-break slideline. Also, thermal conduction and frictional boundary heating can be handled in a straightforward manner. However, other factors, such as the location of the free surface of the deformed chip, are handled better by NIKE2D. The NCSU team is examining the best approach to the overall problem, including the joint use of both Eulerian and Lagrangian codes in a complementary fashion.

Research Hardware

Work on a two-axis cutting force dynamometer, initiated in FY 85,[1] has been completed. In contrast to the piezoelectric force dynamometer used previously,[2] the capacitance gauge sensors can measure steady-state force components. The unit is made of super-invar for low thermal sensitivity, and has a stiffness of approximately 10^6 lb/in. or 2 gf/Å. Since the noise level corresponds to a force of about 2 gf at the tool tip, the capacitance gauge sensitivity is about 1 Å. A technical paper on the dynamometer design and performance is planned for FY 87.

Near the end of FY 86, the Materials Fabrication Division bought an ISI Model DS-130 SEM. It is being installed in Bldg. 691 in support of the chip science effort. The availability of a modern dedicated SEM will be a major benefit, and will allow the future development of new capabilities, such as an internal machining stage for *in situ* study of the chip formation process at high magnification.

Summary and Conclusions

We have completed an extended wear test on the diamond turning of electroless nickel, using an optimum eNi preparation and two tools with different IR absorption characteristics that we had already studied in FY 85. The previous study showed a nearly uniform surface roughness over a cutting distance of 10 000 ft, of which the initial few thousand feet were unavailable for study because of repeated cuts in the test sequence. By including the initial region in the present study, we found that initial wear causes a significant shape change in the tool edge signature in the first 10 000 ft, increasing the machined surface roughness from 5 to 30 Å rms. Thereafter the roughness amplitude remains in a band of 30–50 Å rms over the next 40 000 to 50 000 ft. If we did not know about the initial rise, we would probably assume that the initial irregularity in the tool edge resulted from the sharpening process. Additional insight was also obtained by introducing a tool edge replication technique, which showed a continuing edge recession. This showed that a plateau appearing in the graph of surface roughness vs distance did not imply a lack of continuing diamond loss. Also, the replicas showed the formation of a feed-spaced periodic groove structure that strongly suggests a burnishing interaction between the diamond tool and the workpiece surface. The burnishing evidence indicates the need for a reexamination of the traditional "clean cutting" model of diamond turning, raising questions, for example, about the minimum chip thickness that can be achieved by the diamond tool edge as a function of cutting distance. The IR absorption data that we collected helped us to correctly predict which tool should exhibit the lower wear rate and the larger tendency toward brittle fracture.

The orthogonal cutting tests and the associated numerical modeling of chip flow for a well-characterized aluminum material gave encouraging agreement, indicating that routinely available material test data is a sufficient basis for analysis (at least for low-speed cutting where temperature-dependent properties are not required). This view is further supported by the connection established between the J Integral (an established materials testing concept) and the effective plastic strain-to-failure used as the chip separation criterion in NIKE2D. The NCSU investigation into the complementary use of Lagrangian and Eulerian codes shows promise as a means of reducing the various limitations of either code.

Future Work

Two areas of investigation will be the principal focus for FY 87. The first is the problem

of the minimum attainable depth of cut for a given cutting tool, and its implications in terms of surface roughness and dimensional accuracy. To investigate diamond tools, we have started to collaborate with Prof. N. Ikawa of Osaka University. Since the minimum depth of cut for diamond tools may be only a few nanometers, the experiments are very challenging. Additionally, non-diamond tools with edge radii several orders of magnitude larger will be tested, both because the experimental difficulties are less severe and because the results are of practical interest for the precision machining of non-diamond-turnable materials. The second area of investigation will be the diamond turning of silicon. From an application viewpoint, silicon is of interest for the next generation of liquid-cooled high-energy laser mirrors, because of its low thermal expansion rate and high thermal conductivity. From the viewpoint of increased understanding, silicon is interesting because it is an extensively characterized material, and is available in single-crystal, polycrystalline, and amorphous forms. The limited available evidence indicates that silicon is nominally diamond-turnable, producing a specular finish, but that tool wear is an issue. A transition from ductile, continuous chips to brittle fracture has been reported[10] to occur at depths of cut exceeding 1 μm for single-crystal silicon; a combination of fracture mechanics, code work, and experiments may provide both an understanding of this transition point and a means to shift it to deeper cuts.

Outside Contacts

A faculty-student group headed by Prof. J. Strenkowski is under contract to LLNL to do numerical modeling work at the North Carolina State University Precision Engineering Laboratory. The faculty of the Precision Engineering Laboratory visited LLNL in July, 1986, and gave a series of seminars covering the full spectrum of research at their laboratory. Other contractors include professors A. Mukherjee of UC Davis and I. Finnie of UC Berkeley, who continue to consult on materials science subjects. Interaction on the study of diamond tool wear also continues with the China Lake Naval Weapons Center. Prof. N. Ikawa and Dr. S. Shimada of Osaka University visited LLNL for several days in May, 1986, for initial planning of an Osaka/LLNL collaboration on determining the minimum depth of cut that is attainable with diamond tools. Osaka is providing specially selected diamond tools and edge sharpening techniques, while LLNL is providing cutting tests and characterization studies. Travel grants have been obtained by LLNL from the National Science Foundation and from its Japanese equivalent by Osaka to support this collaboration. During FY 86 members of the LLNL chip science group have also attended the SPIE Conference in San Diego, the Congress of Applied Mechanics in Austin, the ASME Winter Annual Meeting, the SME Precision Engineering Workshop in Boston, and the CIRP General Assembly in Israel; presentations were given at four of these conferences and information exchanged with others interested in chip science. Similar exchanges have taken place with numerous visitors to LLNL.

Financial Status

We are receiving support from several internal and external sources for chip science studies. The total funding for FY 86 was $580 000, of which $400 000 was provided by the Engineering Research Program.

Acknowledgments

The authors would like to acknowledge the contributions and support of many staff members of LLNL: J. Chrislock, J. Dini, J. Ferreira, D. Grayson, A. Madeiros, S. Patterson, K. Shelley, and I. Stowers of ME/MFD; R. Kershaw of ME/MFED; D. Hopkins, R. Rauscher and R. Schenz of EE/NESD; D. Owens of Computations/ASD; and R. McClure and T. Saito of PEP.

1. R. R. Donaldson, C. K. Syn, J. S. Taylor, and R. A. Riddle, "Chip Science: Basic Study of the Single-Point Cutting Process," *Engineering Research Annual Report FY 85*, Lawrence Livermore

National Laboratory, Livermore, CA, UCID-19323-85-2 (1986).
2. J. S. Taylor, C. K. Syn, T. T. Saito, and R. R. Donaldson, "Surface Finish Measurements of Diamond-Turned Electroless-Nickel-Plated Mirrors," *Optical Engineering* 25(9), 1013–1020 (1986).
3. C. K. Syn, J. S. Taylor, and R. R. Donaldson, "Diamond Tool Wear vs Cutting Distance on Electroless Nickel Mirrors," SPIE 30th Annual Symposium, San Diego, CA, Aug. 17–22, 1986, paper No. 676-26 (in press).
4. R. R. Donaldson and D. C. Thompson, "Design and Performance of a Small Precision CNC Turning Machine," *CIRP Annals* 35(1), 373–376 (1986).
5. M. C. Shaw, *Metal Cutting Principles*, (Oxford, Clarendon Press, 1984), pp. 241–244.
6. A. G. Thornton and J. Wilks, "Tool Wear and Solid State Reactions During Machining," *Wear* 53, 165 (1979).
7. R. A. Riddle, "The Effect of the Failure Criterion in the Numerical Modeling of Orthogonal Metal Cutting," 10th U.S. Congress of Applied Mechanics, Austin, TX, June 16–20, 1986, UCRL-93745-ABS (1986).
8. R. A. Riddle, R. D. Streit, and I. Finnie, "The Failure of Aluminum Compact Shear Specimens under Mixed Mode Loading," Fracture Mechanics 18th National Symposium, Boulder, CO, June 24–27, 1985, UCRL-91409-ABS (1986).
9. G. R. Johnson, "Dynamic Analysis of a Torsion Test Specimen Including Heat Conduction and Plastic Flow," *ASME Journal of Engineering Materials and Technology* 103, 201–206 (1981).
10. T. Sugita, K. Ueda, K. Endo, and A. Kishimoto, "High Speed Microcutting of Ceramics: Report I—Basic Study on Glass and Si," *Proc. Japan Soc. Precision. Eng.*, paper No. 119 (1984).

Topometric Interferometry

J. A. Monjes, C. H. Gillespie, and D. F. Edwards
Fabrication Technology Section
Materials Fabrication Division

We have attained the initial goal of this project, which is to characterize convex spherical surfaces accurately, quickly, and reliably by applying the principles of topometric interferometry. A topometric interferometer can provide graphical and numerical information on surface height over an extended area. This article describes the design, construction, and test of a prototype instrument for inspection of spherical surfaces with diameters between 1 and 15.5 mm.

Introduction

The genesis of our project was the need for increased accuracy in characterizing optical surfaces and other components for nuclear weapons programs, the laser program, and thin-film measurement. Our project meets this need through the use of a single instrument that combines multiple interference and wavelength scanning. The instrument, which is extremely accurate and simple to use, can scan an entire sphere in a few minutes. By incorporating automatic rotation of the sphere, scanning time could be reduced to a small fraction of a second.

The topometric interferometer we describe here uses a very narrow bandwidth for the illuminating light. We measure height differences on an object's surface by determining the differential wavelength at which interference fringes are produced. Since two or three fringes are not required in the field of view to gauge fringe distortion, the theoretical fringe spacing can in principle be made arbitrarily large. If there is only one fringe for each wavelength in the field of view, and the wavelength difference is small compared to the separation, the fringe center can be located with respect to the next fringe boundary with greater accuracy. The theoretical limit of the method is a function of the $\delta\lambda$ of the source; a practical limit exists, which is related to the quality of the surface under test. Increasing the test object's quality and reflectivity sharpens the fringe boundaries and makes the measurement more accurate.

We invested considerable thought and effort in the investigation, construction, and testing stages of the project. During the preliminary phase we undertook extensive research into the literature as well as analytical work to determine the best technique for achieving a resolution of 1 to 2 μm and an accuracy of 0.001 μm or better. This project is indebted to the pioneering research of many investigators, especially Tolansky[1] on the characterization of metallic surfaces; Herriott,[2] Heintze,[3] Gates,[4] and Biddles[5] on the characterization of spherical surfaces; and Pilston and Steinberg[6] and others on the use of discrete wavelength intervals to reduce fringe spacing.

We considered several techniques, including scatterometry, profilometry, and interferometry. In topometric interferometry, an extended area is plotted simultaneously by measuring contour lines of equal height. By contrast, a profilometer plots a single line of varying levels along a pre-established path, and scatterometry explores the surface point by point. We concluded that topometric interferometry offered the best chance of achieving maximum accuracy, resolution, and speed, along with simple data interpretation.

Next came experimental verification of our theoretical studies. We completed the layout of an experimental topometric interferometer and purchased the main components for the instrument. We then constructed an experimental setup using a helium-neon (HeNe) laser as a light source. Even with no tuning capability we were able to demonstrate good

fringe visibility and overall mechanical feasibility.

Later phases of the investigation included a study of tunable light sources. We acquired an infrared (IR) diode laser with control unit, and designed and built a system adapted to the IR wavelength. We calibrated the infrared laser wavelength as a function of power output and diode temperature. We also designed and constructed motorized micro-stages for ball positioning, and assembled and aligned the topometric interferometer on new, magnetically suspended bases. We were able to demonstrate the instrument's capability of resolving interference fringes on the recorded interferogram with an accuracy approaching 0.1 Å per millimeter.

Technical Status

Topometric Approach

The elements of a topometric interferometer are shown in Fig. 1. The topometric instrument basically measures surface profile along lines of constant level. From these measurements, statistical data for the entire area can be calculated. The instrument can also provide automatic statistical information by direct measurement, and it can be adapted to measure surface curvature and roughness. Our calculations show that the theoretical limiting accuracy of such an instrument approaches 10^{-10} m.

When the instrument is operating, the test object is illuminated by infrared laser light with an extremely narrow frequency spectrum. The frequency of this light is scanned over a given bandwidth by tuning the source, and as a result the light is made to resonate in the reflecting cavity formed by the surface of the test object and the reflection-coated outer surface of the illumination lens. The cavity resonance produces localized multiple interference fringes, which are detected by the television camera for analog display. The television monitor shows a magnified view of lines of constant phase, representing precise increments of distance between the test surface and the reference (illumination lens) surface.

An entire sphere with a diameter ranging between 1 mm and 1.5 cm can be explored in less than 5 min by rotating it to not more than six to eight successive angular positions. The maximum diameter of the test object is limited by the curvature of the illumination lens, and the minimum diameter

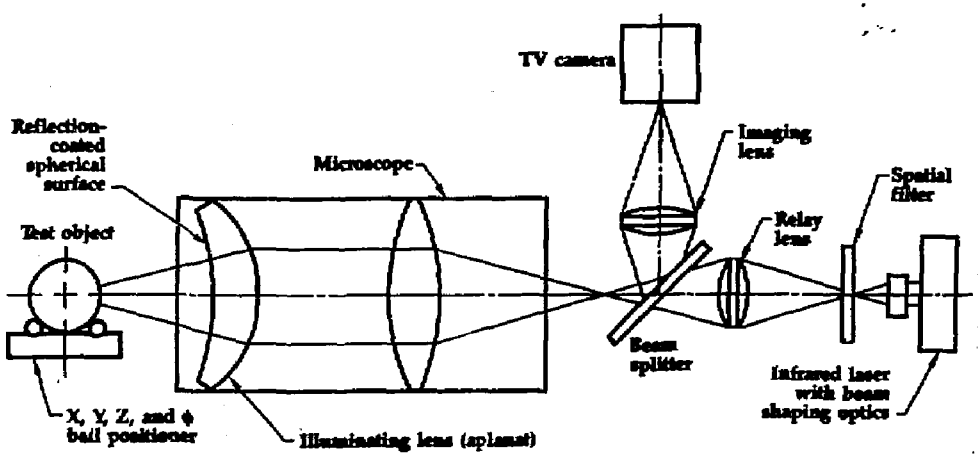


Figure 1. Elements of a topometric interferometer.

of the test object is limited by the desired accuracy and resolution.

Theoretical Considerations

The basic theory and analytical description of this method of measurement have been previously published.[7] The following discussion represents a summary of the relevant points and conclusions.

Accuracy of Measurement. The accuracy limit is proportional to the line width of the illumination beam. That is,

$$\delta h = (M/2)\delta\lambda \quad , \tag{1}$$

where M is the number of half-wavelengths in the space between the object's surface and the reflecting surface of the illuminating lens.

Tunability Range of the Probing Beam. To choose a light source, we must first define the tunability range of the probing beam. The topometric interferometer detects lines of constant level by scanning the wavelength of the source. For this reason, the total differential height δh to be explored is the factor that determines the tunability range $\Delta\lambda^*$.

Here we can define two cases in which the differential height δh fulfills the condition $\delta h < \lambda/2$ or the condition $\delta h > \lambda/2$. In both cases we can find the tunability range by considering the Fabry-Perot resonant condition established between the object and the reflective surface of the lens. This condition is discussed in a previous publication,[8] where the smallest resolvable height differential δh is shown to be

$$d_1 - d_2 = \delta h = (M/2)\,\delta\lambda \quad . \tag{2}$$

Since

$$M/2 = d/\lambda \quad , \tag{3}$$

we have

$$\delta h = (d_1/\lambda_1)\,\delta\lambda \quad , \tag{4}$$

where d_1 and d_2 are the variable spacing between points 1 and 2, and M is the integral number of half-wavelengths in the distance d between the object surface and the illuminating lens.

In the first case, $\delta h < \lambda/2$. Therefore, M is a constant. We have previously shown that

$$\Delta\lambda^* = (\Delta h/d_1)\,\lambda_1 \quad . \tag{5}$$

In the second case, $h > \lambda/2$. Therefore, M_1 is not equal to M_2, and we have $M_2 - M_1 = n$. Therefore,

$$\Delta\lambda^* = \frac{2\Delta h - n\lambda_1}{2d_1 + n\lambda_1} \times \lambda_1 \quad . \tag{6}$$

Here we see that when $M_1 = M_2$ we have $n = 0$, and Eq. (6) reduces to Eq. (5). Also,

$$\frac{2\Delta h - n\lambda_1}{2d_1 + n\lambda_1} < \Delta h / d_1 \quad . \tag{7}$$

Therefore, the tunability range defined by Eq. (5) is larger than the one given by Eq. (6), and we can safely use Eq. (5) to determine $\Delta\lambda^*$. By combining Eqs. (4) and (5), the selection of $\Delta\lambda^*$ becomes independent of d. Therefore

$$\Delta\lambda^* = \Delta h\delta\lambda/\delta h \quad , \tag{8}$$

where the tunability range is a function of the desired range of measurements Δh, the line width of the probing beam $\delta\lambda$, and the desired accuracy δh.

Fabrication Technology

The equations given above show that for constant Δh, δh, and $\Delta\lambda^*$, a longer wavelength has the advantage of increasing the working distance d, allowing the measurement of spheres with a wider range of diameters while using the same illuminating lens and obtaining the same accuracy. Use of a longer wavelength also reduces the effects of mechanical movement and thermal variation on the structure. If infrared light is used, a visible-light alignment laser can be incorporated into the design.

The most important conclusion of our previous study is that the tunability range of the source needs to be only of the order of 10 Å. Therefore broad-band incoherent sources are not required, and we need consider only

laser diodes and dye lasers. For reasons of availability and cost, we selected an infrared diode laser with a wavelength range of 0.7 to 0.8 μm and a control unit that provides for temperature-tunable wavelength.

The Infrared System

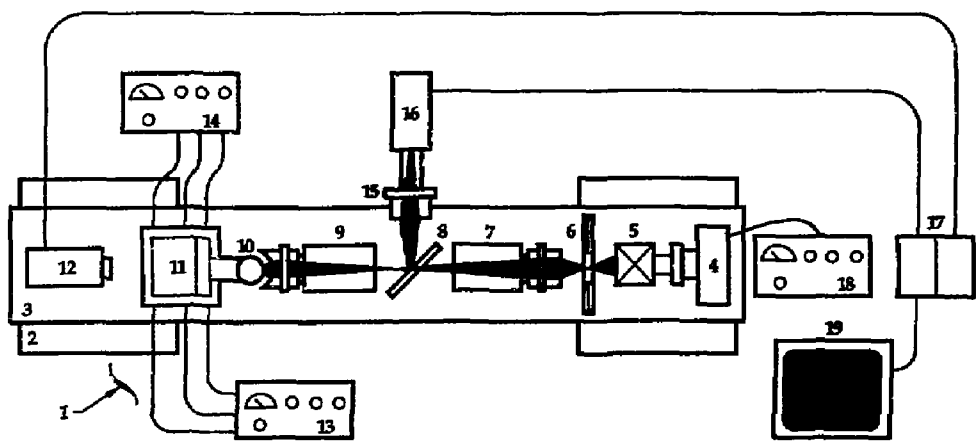
Our present light source is a Liconix Diolite 800 system, with a maximum power of 10 mW, wavelength of 780 ± 20 nm, bandwidth of less than 20 MHz or 0.004 Å, and wavefront distortion of less than λ/4. The laser diode is a Hitachi 7802E operating in a single longitudinal mode (TE_{00}). The control unit has a sensitivity of 1 mW/V, a temperature control of 10°C/V with 0.1°C steps, a wavelength control of 0.05 nm, and a tunability range of ± 20 nm.

We have calculated that the required tunability range is $\Delta\lambda^* = 10$ Å. The Diolite 800 has a range of $\Delta\lambda^* = 400$ Å. This offers the possibility of reducing the tunability range by a factor of 40, with an improvement in wavelength control accuracy from a $\delta\lambda$ of 0.5 Å to 0.0125 Å.

The specifications discussed above are those given by the manufacturer. For the actual Diolite 800 unit we used, our calibration (Table 2) gave slightly different values, with wavelength control of 0.18 Å, tunability range of 70 Å, and a tunability constant of 1.8 Å/°C for output power less than 5 mW and diode temperatures between 7°C and 20°C. The experimental results presented here are based on these actual calibration results and not the manufacturer's specifications.

The Topometric Interferometer Setup

The layout of the topometric interferometer is shown in Fig. 2, which indicates the essential simplicity of the instrument. Figures 1 and 2 show the optical path. Figure 3 shows the entire setup as we assembled it in the laboratory, Fig. 4 is a closer view of the object test stand and illumination lens, and Fig. 5 shows the opposite end of the instrument with infrared laser and focusing microscope. Figure 6 is a semi-schematic view showing the test object and the focusing mechanics of the illuminating lens.



1. Air-suspended optical table
2. Magnetic vibration isolators
3. Optical bench
4. Laser and beam-shaping optics
5. Beam attenuator
6. Spatial filter
7. Relay optics
8. Beam splitter
9. Aplanatic microscope
10. Test object
11. Ball centering stages
12. TV alignment camera
13. Remote control for micrometric stages
14. Remote control for piezoelectric stages
15. Fringe-imaging lens
16. Fringe-imaging TV camera
17. Power supply for TV camera
18. Control unit for infrared laser
19. TV monitor

Figure 2. Layout of experimental topometric interferometer.

Centering the test sphere on the axis of the illuminating lens is greatly facilitated by the use of remotely controlled motor micrometers. A similar remote control operates piezoelectric positioners for fine centering. The overall stability of the system is greatly improved by the use of a magnetic vibration isolator in addition to the air-isolated optical table.

We used a Leitz periplan microscope objective No. 559141 with NA of 0.22 and focal length of 16 mm. By introducing a specially designed aplanat illuminating lens in front of this objective we obtain a NA of 0.33, which will cover a circular area of 4.7-mm diameter on a test sphere of 15-mm diameter. For the measurements reported here the outer surface of the aplanat was made only 50% reflective instead of the optimum 90%. This compromise was made to facilitate alignment at the expense of resolution.

Test Procedure

We use the following equivalent of Eq. (2) to define the minimum resolvable height differential $\delta h = (M/2)\,\delta\lambda$. This in turn permits

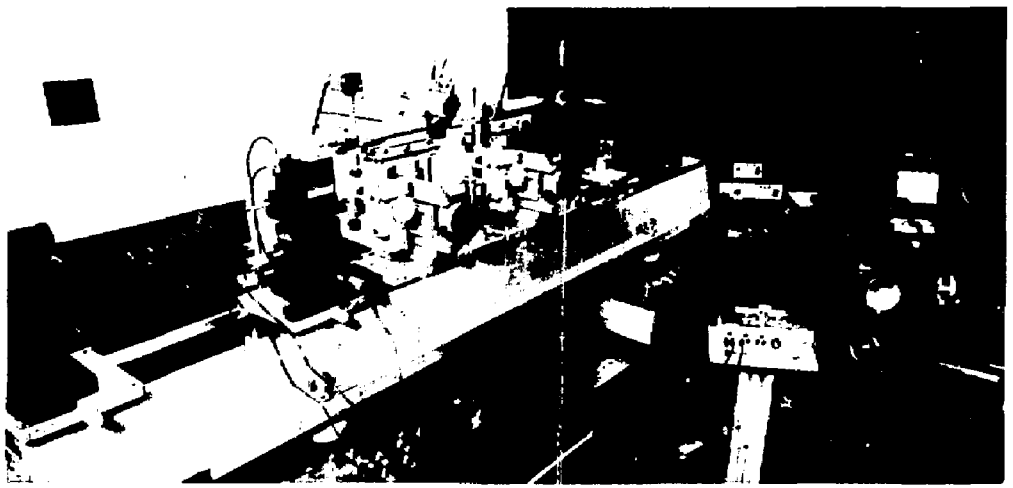


Figure 3. Overview of laboratory setup.

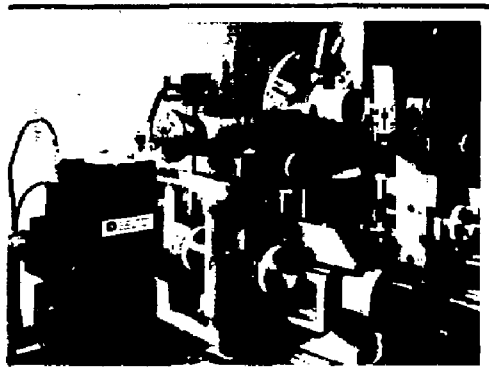


Figure 4. Closeup of object test stand and illumination lens.

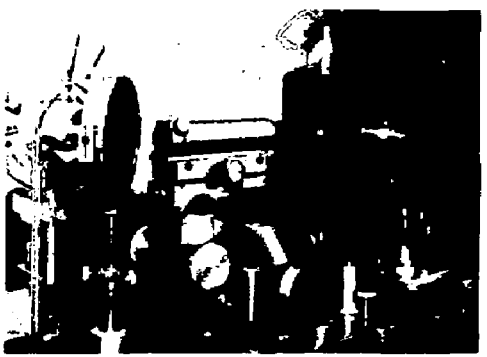


Figure 5. Closeup of infrared laser and focusing microscope.

the evaluation of the total height differential Δh between two areas of the test object. The equivalent expression is

$$\Delta h = (M/2)\,\Delta\lambda \quad , \tag{9}$$

where Δh is the total height differential between two areas, and $\Delta\lambda$ is the fringe displacement as measured on the interferogram.

In the present case, $\Delta\lambda$ is not calibrated by measuring the fringe separation at $\lambda/2$, but by measuring $\Delta\lambda = \lambda_1 - \lambda_2$ as the laser diode is being tuned. We must therefore calibrate the laser wavelength as a function of the diode temperature, which is regulated by the laser control unit. The required calibration values as functions of power and temperature are given in Tables 1 and 2. Graphical representations of these data are given in Figs. 7 and 8. We can see that the laser operates at $\lambda = 784.9$ nm from zero power up to 2.7 mW. At 2.9 mW the laser "jumps mode" and the output changes to $\lambda = 786.8$ nm.

Wavelength variation as a function of diode temperature at a constant output of 5 mW is shown in Table 2 and Fig. 8. We observe that at 5 mW and 20°C there are two possible values of λ. We therefore chose to operate at temperatures less than 20°C and power levels less than 5 mW during our experiments. Within this range we found that the calibration factor or $\Delta\lambda/\Delta T$ is 1.8 Å/°C.

Having obtained these calibration factors we proceeded to make measurements on a lapped spherical ball of 13.81-mm diameter. This ball was partially gold-plated with a plating thickness varying between 0.4 μm and zero. A zone approximately 2.5-mm wide along the equator of the ball was left unplated, creating a step with a height varying between 0.4 μm and zero as shown in Fig. 9. The ball was centered with respect to the illuminating lens, with the upper edge of the step approximately on the centerline of the beam. We obtained the television monitor interferogram shown in Fig. 10 with the ball displaced from the concentric position approximately 1.5 λ closer to the lens.

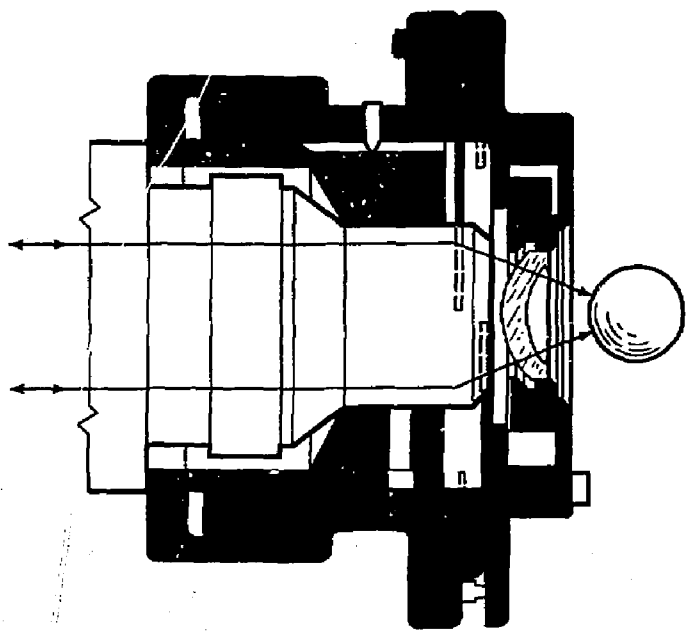


Figure 6. Semi-schematic view of test object and illuminating lens.

The laser was stabilized at 15°C and 2.5 mW, which corresponds to $\lambda_0 = 784.5$ nm. Figure 11 shows an interferogram of the same area with the laser setting unchanged and the ball less than 0.5 λ from the concentric position.

Then the laser temperature was reduced to 14°C. From the calibration curve we obtain $\lambda_1 = \lambda_0 - 1.8$ Å, so $\Delta\lambda = \lambda_0 - \lambda_1 = 1.8$ Å.

Figure 12 is an interferogram at an intermediate temperature, taken while the fringe of Fig. 12 is scanning the ball from the edge of the field toward the center.

Figure 13 shows the stabilized fringe when the laser has achieved $\lambda_1 = 14$°C. By reducing the temperature still further, the fringe is made to converge on the center, while a new fringe is generated at the edge of the field of

Table 1. Laser wavelength vs output power (T = 20°C).

Power (mW)	Wavelength (nm)
0.7	784 9
1.0	784.9
2.0	784.9
2.7	784.9
2.9	786.8
3.0	786.8
5.0	786.8
7.0	786.8
9.9	786.8

Table 2. Laser wavelength vs temperature (Power = 5 mW).

Temperature °C	Wavelength (nm)
7.4	783.1
10.0	783.6
15.0	784.5
20.0	786.8
25.0	787.3
30.0	788.6
35.0	790.1
38.0	791.0

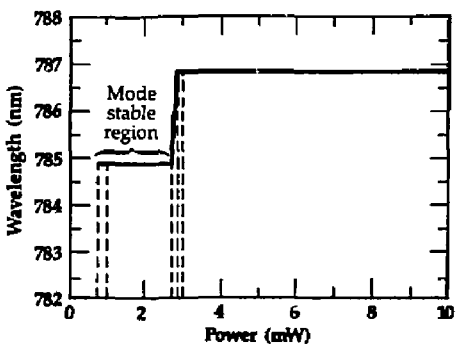


Figure 7. Laser output wavelength vs output power.

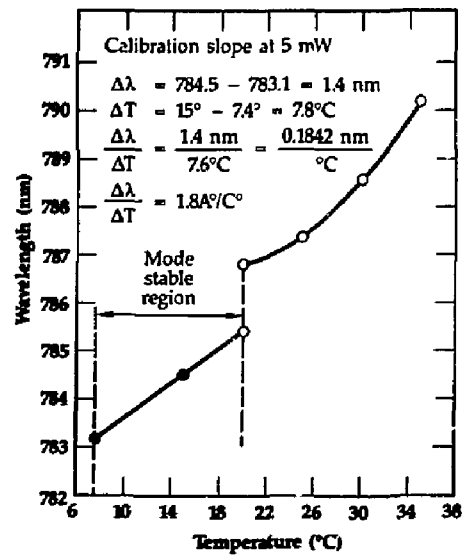


Figure 8. Laser output wavelength vs temperature.

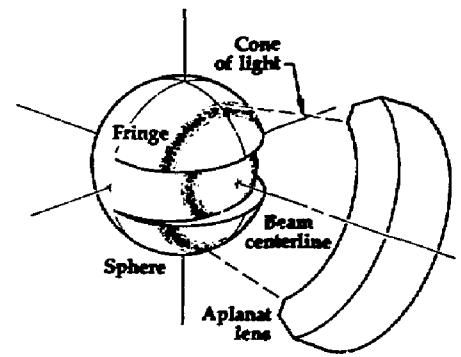


Figure 9. Fringe pattern on target with recessed zone.

view. These three images show how tuning the laser at $\Delta\lambda$ intervals can make a single fringe scan a field approximately 5 mm across in less than 10 s. The actual scanning time is less than 1 s, but the temperature oscillates for a few seconds before stabilizing at the new value.

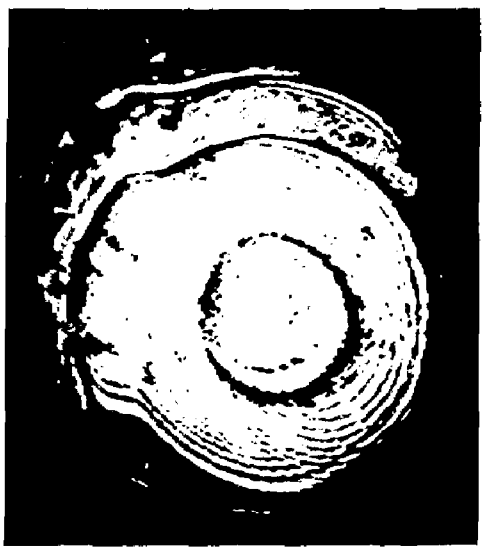


Figure 10. Interferogram with target 1.5 λ from concentric position.

Figure 12. Interferogram showing fringe scanning ball as laser temperature is changing from 15° to 14°C.

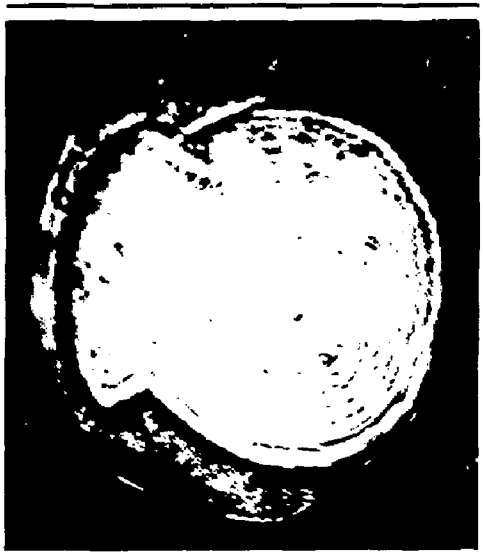


Figure 11. Interferogram with target less than 0.5 λ from concentric position.

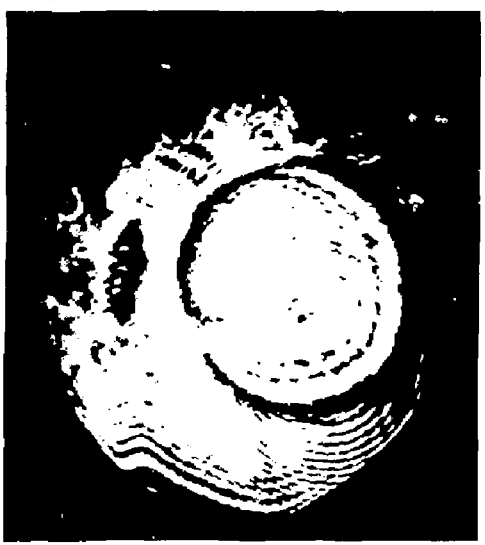


Figure 13. Interferogram showing fringe stabilized after laser attains 14°C.

Interferogram Evaluation

Continuing with the test object described above, we can evaluate the thickness of the gold film by measuring the fringe discontinuity on the interferogram at the area of discontinuity. We have shown that differential heights on the object surface are given by

$$\Delta h = (M/2)\,\Delta\lambda \quad , \tag{10}$$

where M is the number of half-wavelengths in the distance d between the object surface and the front of the aplanatic illuminating lens. Thus, in terms of full wavelengths, we can obtain

$$\Delta h = M_0\,\Delta\lambda \quad , \tag{11}$$

where M_0 is the number of full wavelengths in the distance d. The distance between the test object and the aplanat is

$$d = (D_0/2) - R_a$$

where D_0 in this case is 13.810 mm, the diameter of the sphere, and R_a is 7.830 mm, the surface radius of curvature of the aplanat. These figures give a value of 0.925 mm for d. From the calibration curve (Fig. 8) the value for λ at 15°C is 784.5 nm, so M_0 is d/λ, which equals 1179.

Next we determine the scale of the interferogram. We know that the $\Delta\lambda$ between Figs. 10 and 12 is the value corresponding to a ΔT of 1°C. Applying this value to the calibration curve gives a $\Delta\lambda$ of 1.8 Å. By measuring the diameter differential between the central fringes on the same figures, we obtain a ΔD_f of 20 mm. The interferogram scale is therefore

$$S_1 = \frac{W_g}{\Delta D_f} = 0.09 \text{ Å / mm} \quad .$$

The fringe discontinuity in Fig. 10 for the left side (Δf_l) equals 9.3 mm, and for the right side (Δf_r) equals 8.0 mm. The fringe shift for each side is therefore

$$\Delta\lambda_l = \Delta f_l \times S_1 = 0.837 \text{ Å} \quad ,$$

and

$$\Delta\lambda_r = \Delta f_r \times S_1 = 0.720 \text{ Å} \quad .$$

Since the height differential is given by

$$\Delta h = M\,\Delta\lambda \quad , \tag{10}$$

we obtain

$$\Delta h_l = 1179 \times 0.837 = 987 \text{ Å} \quad ,$$

and

$$\Delta h_r = 1179 \times 0.720 = 849 \text{ Å} \quad .$$

The previous values can be verified by the interferogram of Fig. 9, which shows the spacing between three fringes. We should note that only the two central fringes can be used to evaluate the interferogram because these fringes are on a spherical surface projected onto a plane. The marginal fringes will therefore appear to be spaced more closely than they actually are. From the two central fringes, therefore, we obtain

$\lambda/2 = 3923$ Å $\equiv 16$ mm on the interferogram.

The interferogram scale has the following value:

$$S_2 = 3923/16 = 245 \text{ Å/mm} \quad .$$

The fringe discontinuities as measured on the interferogram have the values for Δf_l of 4.5 mm and for Δf_r of 4.0 mm. Thus the differential height values are

$$\Delta h = \Delta f \times S_2 \quad ,$$

$$\Delta h_l = 1100 \text{ Å} \quad ,$$

$$\Delta h_r = 980 \text{ Å} \quad .$$

As expected, due to spherical curvature error, these values are larger than the ones obtained by fringe scanning. We obtain a discrepancy of approximately 120 Å, or $\lambda/65$, which is a good value for reading $\lambda/2$ fringe spacing.

Accuracy could be greatly improved by using a higher reflectivity than the value of 50% we selected for the aplanat's surface, by using a longer wavelength, by decreasing the distance between object and illuminating lens, and by increasing the frequency stability of the light source.

System Errors

Let us assume that we can measure fringe location with an accuracy of ± 0.2 mm either by reading λ/2 fringe spacing or by fringe scanning. The errors for λ/2 fringe spacing would then be equivalent to an interferogram scale error (S_2) of 245 Å/mm, a reading error (δ_r) of 0.2 mm, and a surface measurement error (δl_2) of $S_2 \times \delta_r$ or ± 50 Å. For fringe scanning, the corresponding errors would be an interferogram scale error (S_1) of 0.09 Å/mm, a reading error (δ_s) of 0.20 mm, and an interferogram error (δ_λ) of $S_1 \times \delta_s$ or ± 0.018 Å.

According to Eq. (10) this interferogram error translates into an actual error on the surface of the sphere as given by

$$\delta_R = M_0\, \delta\lambda \quad .$$

In the present case we have a large value for M_1, which equals 1179, and the error may be computed as

$$\delta_{R1} = 1179 \times 0.018 \text{ Å} = 21 \text{ Å} \quad .$$

For more accurate measurements, the value of M should be reduced to a few wavelengths; that is,

$$10 < M < 100\ S_0 \text{ for } M = 50 \quad .$$

So if all other conditions remain the same, we have

$$\delta R_0 = 50 \times 0.018 \text{ Å} = 0.9 \text{ Å} \quad .$$

These error calculations confirm our theoretical predictions concerning the performance of the topometric interferometer. We conclude that for maximum accuracy the spacing between lens and object should be minimized. It is also plain that longer wavelengths are desirable to reduce the value of M and permit measurement of a large range of object sizes with maximum accuracy. We further conclude that higher fineness is required to improve fringe location to a value better than 0.2 mm; this would require a mirror reflectivity greater than the value of 50% used in our experiments.

Future Work

The advantages of using multiple interference fringes in interferometry have already been shown.[1–6] Our investigation was designed to demonstrate the benefits of fringe scanning as combined with multiple interference interferometry.

We have evaluated both methods in this project through interpretation of interferograms of multiple interference at constant wavelength (e.g., Fig. 9) as well as interferograms of multiple interference with wavelength tuning (e.g., Figs. 10–12). Our results show that fringe scanning provides greater accuracy because of increased fringe spacing for small Δλ increments.

Several other important benefits are associated with the fringe scanning technique. For example, all points of the surface under measurement are scanned in a continuous manner, without the untested areas characteristic of techniques using fixed wavelength and narrow fixed fringes. Additionally, the use of longer wavelengths actually increases the accuracy of the technique and its adaptability to different object diameters [Eq. (4)]. And finally, the technique permits simultaneous fringe presentation over an extended area, offering the possibility of automatic, synchronous fringe scanning and ball rotation for high production rate. These last two points are more appropriate for funding under a programmatic development or with industrial participation, since they are highly relevant to industrial applications such as high-speed, automated inspection of ball and roller bearings.

Financial Status

Total expenditure for this project was $187 000, of which $133 000 was for manpower and manufacturing and $54 000 was for procurement.

1. S. Tolansky, *Multiple-Beam Interference Microscopy of Metals* (Academic Press, London, 1970).
2. D. R. Herriott, "Multiple Wavelength-Multiple Beam Interferometry," *J. Opt. Soc. Am.* **51**, 1142 (1961).
3. L. R. Heintze, "A Multiple-Beam Interferometer for Spherical Wavefronts," *Appl. Op.* **6**, 1924 (1967).
4. J. W. Gates, "An Interferometer for Testing Sphericity," *Optics in Metrology*, P. Mollet, Ed. (Pergamon Press, New York, 1960).
5. B. J. Biddles, "A Non-Contracting Interferometer for Curved Surfaces," *Op. Acta* **16**, 137 (1969).
6. R. G. Pilston and G. Steinberg, "Multiple Wavelength Interferometry with Tunable Source," *App. Op.* 8(3), 553 (1969).
7. J. A. Monjes, C. H. Gillespie, and D. F. Edwards, "Topometric Interferometry," *Engineering Research Progress Report FY 84*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-84-3 (1984).
8. J. A. Monjes, C. H. Gillespie, and D. F. Edwards, "Topometric Interferometry," *Engineering Research Progress Report October 1983–March 1984*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-84-2 (1984).

Advanced Fabrication of Optical Materials

P. P. Hed and K. L. Blaedel
Fabrication Technology Section
Materials Fabrication Division

The fabrication of high-precision optical elements for new generations of high-power lasers requires a deterministic method of generating precision optical surfaces entailing considerably less time, skill, and money than present optical techniques. Such a process would use precision computer-controlled machinery with ongoing *in situ* metrology to generate precise optical surfaces. The implementation of deterministic processes requires a better understanding of the glass-grinding process, especially the control of ductile material removal. This project is intended to develop the basic knowledge needed to implement a computer-controlled optics-manufacturing methodology.

Introduction

In its July, 1985 report, the Advanced Laser Development Research and Requirements Committee stated that advanced optical fabrication techniques were an engineering requirement of the highest priority for large, high-power laser applications. The Committee found that future laser developments would be seriously limited unless optical fabrication costs were substantially reduced through new manufacturing technology, specifically deterministic processes that would markedly decrease production time and simultaneously compensate for an otherwise serious shortage of skilled craftsmen. Such processes would use precision machinery under computer control to generate optical surfaces to exacting tolerances and would subsequently verify surface quality through *in situ* metrology. An automated machining technology with accurately predictable performance could be expected to perform consistently within acceptable tolerance limits.

Great strides have been made in automated machining technology during the past thirty years. Modern structural and thermal engineering design methods have helped create unusually stable machine tool platforms. Computer numerical controls in conjunction with meticulously prepared diamond cutting tools have resulted in steadily improving dimensional accuracies in surfaces generated through turning processes. It is now possible to diamond-turn selected parts of more than 1-m-diam to mirror-like surface finishes, with figure tolerances down to a few tens of nanometers, and with surface roughnesses of under 100-Å rms.

However, these superior finishes can be attained only with certain metals such as electroless nickel, copper, and perhaps aluminum. Glasslike materials, often preferred by optical system designers for their dimensional stability, are not so well suited to diamond turning. Diamond turning is essentially a crushing process when applied to brittle materials such as glass, and the resulting surface finish is seriously impaired by a damaged layer up to 50–100 μm deep caused by the crushing forces. Moreover, short wavelength optical applications require figure accuracies that are difficult to attain even in metallic materials.

It is plain that the development of a process for the high-speed fabrication of large optical components requires more than just a high material removal rate. The primary requirement is a reduction in the total time devoted to the fabrication and certification of the optical element. This beginning-to-end fabrication process includes initial part setup, material removal processing, in-process and final inspections, and any handling, transporting, and setup necessary during the overall manufacturing cycle.

Cost reductions may be possible at any given step in the process. However, the largest reduction would result by taking a comprehensive look at the process as a

whole. This means examining all the steps individually and also searching for ways to integrate the sequence of manufacturing operations more efficiently. Reducing the number of operations would be particularly valuable. One operation that might be eliminated is the extremely time-consuming transportation and setup now required whenever the part is inspected after material removal.

We believe that there are two approaches to reducing the cost of producing advanced laser systems. One approach focuses on the science and mechanics associated with individual manufacturing operations. The other approach, while drawing on the results of the first, explores a new method of integrating the manufacturing steps. An example of this latter approach would be a new machine tool that uses a deterministic material removal process and simultaneously incorporates an *in situ* metrology capable of near-real-time tracking of the surface being generated.

Let us first briefly consider the individual manufacturing steps, in particular the material removal steps. The manufacture of high-quality glass optics has traditionally involved three quite different material removal stages. The first stage is a roughing or grinding stage in which the optical blank is machined to approximate final shape by an abrasive bound into a tool, for example, a grinding wheel. After completion of this stage the workpiece has a good overall contour, but the surface is poor, manifesting considerable roughness accompanied by subsurface damage. This stage of manufacture has been the central focus of our continuing research. We will refer to it as "bound abrasive grinding."

At some point the workpiece may be transferred to the second manufacturing stage, lapping, which is generally characterized by a lower material removal rate. During lapping the workpiece is machined by an abrasive that is free to circulate in a slurry between the tool and the workpiece. Whether lapped or not, the work eventually passes to the final stage, polishing. This stage, which has the slowest material removal rate of all, is a low-force variant of bound abrasive grinding in which very fine abrasive is held in a soft matrix such as pitch, and is augmented with a slurry, which has both chemical and abrasive qualities. These two latter stages, lapping and polishing, remove the roughness and subsurface damage caused by the initial bound-abrasive grinding. However, the metrology required to maintain or improve the surface contour of the workpiece during these stages is extremely costly and time-consuming. This is especially true when the initial bound-abrasive grinding operation has produced a substantial damage layer that must subsequently be removed.

It is important to note that polishing is not a deterministic process. The compliance of the pitch precludes any accurate estimate of the location of the machined surface. The longer polishing proceeds beyond a surface measurement, the more uncertain the actual surface contour becomes. Thus a large number of time-consuming measurements are routine during polishing.

The second and third material removal stages of manufacture are also of concern in the overall fabrication process, but they are not part of our investigation. In the FY 85 *Engineering Research Annual Report*,[1] we present some insight into these processes. However, in FY 86 we have focused on the grinding step because it is a major source of subsurface damage. This damage is, in turn, a major contributor to the high processing costs and manufacturing complexities associated with the generation of precision optical surfaces.

Our general goal was to demonstrate a precision-grinding fabrication process for optical glass resulting in minimal surface roughness and damage, which in turn minimizes the subsequent lapping and polishing required to finish an optical element. We therefore conducted a series of experiments to study the relationship among assorted bound-abrasive parameters and the associated surface roughness and damage. This aspect of our work was a continuation of our FY 85 Engineering Research Program activities.[2] We simultaneously pursued the design and construction of a precision grinding machine. This is a departure from our FY 85

approach, which focused on diamond crushing. Because of the subsurface damage we observed in association with the diamond crushing process, we now feel that the main role of this method is roughing. We have therefore turned our attention toward precision grinding machinery.

Technical Status

Bound Abrasive Studies

As we stated in our annual report for FY 85,[1] the first objective in our fine grinding study is the establishment of an engineering data base. Such a data base will permit the development of a method to accurately predict the depth of the subsurface damage resulting from the grinding process. Ideally such predictions would be based on an easily measured parameter, which in our case would be the surface roughness of the workpiece. We would then correlate the observed subsurface damage with the associated grinding conditions. Knowledge of this important relationship will be helpful in guiding the development of advanced fabrication processes, and will be immediately useful for process control in current fine grinding practice.

Our plan for building the engineering data base consists of three objectives. The first objective is the precision grinding of glass samples with a single diamond pellet and the measurement of associated surface roughness and subsurface damage. The second objective is the precision grinding of glass samples with multiple diamond pellets (Fig. 1) and continuous wheels, again with measurements of surface roughness and subsurface damage. The third objective is the correlation of grinding parameters with the measured surface roughness and subsurface damage. During FY 85 we focused our attention on the first and third objectives. During FY 87 we plan to pursue the second objective when the appropriate precision grinding machinery is available.

Experimental Preparations

The glass samples[3] we used in our experiments were Schott BK-7, Zerodur, and Corning No. 7940 fused silica measuring 48-mm-diam by 20-mm thick. All surfaces of the samples were carefully polished to remove any traces of subsurface damage

Figure 1. Multiple-pellet diamond grinding tool.

produced during the initial fabrication process. Previous experience[1] has verified that our sample preparation method yields surfaces free of residual damage.

Our diamond grinding experiments were made on the polished glass samples after the surface had been lapped for about 30 min on a cast-iron tool with a 500-g load. We used 9-μm aluminum oxide slurry containing a small amount of sodium citrate.[4] The lapped surface is necessary because diamond grinding is ineffective on polished surfaces. The surface finish of the 9-μm lapped surfaces was 2.5 μm, 2.3 μm, and 2.3 μm peak-to-valley (P-V) for the BK-7, Zerodur, and fused silica, respectively. Previous measurements indicated that the 9-μm lapping as used here produces a damaged layer of 10 μm or less as measured by the taper-polishing technique.[1] Knowledge of this initial damage depth aids us in determining the minimum amount of glass that must be removed in the grinding experiments. We need to be certain that any measured damage was produced during the grinding process and was not left over from a previous lapping or fabrication step.

Single-Pellet Grinding

The single-pellet grinding experiments were performed using the single-axis Pneumo surfacing machine.[5] The glass samples were held in a vacuum chuck in a fixed position and the single-pellet tool was mounted as a fly cutter with 280-mm diameter. For these experiments the tool rotation speed and crossfeed were fixed at 1750 rpm and 5 mm/min. The lubricant was 4.5% Rust-Lick B by volume[6] in deionized water at 22°C. For a given experiment the depth of cut was also fixed. Three different pellet types were used (Table 1). The concentration of the abrasive in the bonding material was 12.5 vol%. The protrusion of the abrasive (Table 1) was determined by two methods. The first method requires making an impression of the dressed pellet using Frater's Facsimilike[7] and then measuring the impression with the Federal No. 2000 Surfanalyzer profilometer.[8] This impression method prevents excessive wear on the 2.5-μm-radius profilometer stylus tip. The second method requires focusing a microscope on the diamond tip and then on the bond, and measuring the difference between the two microscope settings.

Prior to each experiment, the single-pellet tool was dressed with an aluminum oxide stick applying 20 cuts, each 0.5 μm deep, taken at 350 rpm and 25 mm/min crossfeed. Contact between the rotating pellet and sample was determined by observing tool marks on the painted surface of the sample. This method was found to be more reliable than an acoustic method because of the high background noise produced by the lubricant flow. The sample was advanced a given distance, and a uniform layer was removed from the sample on each cut. This cutting step was repeated for a given number of passes. The results for these single-pellet grinding experiments are summarized in Table 2. Shown in the fifth column is σ, the P-V surface finish, representing the arithmetic average and standard deviation of a

Table 1. Properties of pellets.

Grinding application	Abrasive type	Abrasive size (μm)	Abrasive protrusion after dressing (μm)	Bond material	Bond Rockwell hardness (B-Scale)	Pellet diameter (mm)	Source
Single Pellet	Diamond	53–65	7–12	Bronze	109	13	DTI[a]
Single Pellet	Diamond	180–250	25–30	Bronze	90	10	DTI[a]
Single Pellet	B_4C	180–250	25–30	Bronze	90	13	DTI[a]
Multi-Pellet	Diamond	15–25	4.5	Cobalt bronze	98–104	6.4	PDT[b]
Multi-Pellet	Diamond	6–15	1.5–2.0	Phenolic	N/A	6.4	Elgin[c]
Multi-Pellet	Diamond	6–15	1.5–2.0	Cu-Sn-Resin	N/A	6.4	Elgin[c]
Multi-Pellet	Diamond	6–15	1.5–2.0	Cu-Phenolic	N/A	6.4	Elgin[c]
Multi-Pellet	Diamond	6–15	1.5–2.0	Epoxy	N/A	6.4	LLNL

[a] Diamond Technology Industries, Lake Forest, IL
[b] Diamond Precision Tool Co., Elgin, IL
[c] Elgin Diamond Products Co., Elgin, IL

number of readings on similar samples. In a few cases, more than one finish measurement was made on the same sample. The subsurface damage depth, δ, given in the sixth column of Table 2, was determined by the taper polishing method.[1] Here also we give the arithmetic average and standard deviation for a number of measurements on similar samples. The last column, K, gives the ratio of the subsurface damage depth to the surface finish (δ/σ).

The first three rows of Table 2 compare the single-pellet grinding properties of BK-7, Zerodur, and Corning No. 7940 fused silica for 53–65 μm mesh diamond. The Corning No. 7940 fused silica, which has the largest microhardness of the three materials, also has the smoothest finish. Time restrictions did not permit further experiments using the other abrasives with the fused silica. In all cases the Zerodur surface was smoother than the BK-7 surface. This was true for the larger diamond size, the boron carbide, and the greater depth of cut. The 1.2-μm P-V for Zerodur using 180–250-μm diamond is suspect. Previous experience indicates that P-V values increase with increasing abrasive size. For this reason we neglect this value in our analysis.

From Table 2 the δ values for BK-7 (rows 4 and 6) appear to be inconsistent with the notion that the depth of damage increases with depth of cut for a given diamond size. The value of δ in row 6 for BK-7 appears to be too small when compared with the damage depth values in rows 4, 8, and 9. No explanation can be offered at this time. For the BK-7 sample of row 9, three sparking-out passes were made at the end of the experiment. This means that three additional cutting passes were made without advancing the tool. This sparking-out appears to have improved the surface finish (compare rows 8 and 9).

The K-values in the last column show a consistent correlation between the P-V roughness and the depth of damage. As will be shown below, the surface finish for fused silica tends to improve (i.e., smaller P-V) as cutting progresses. We believe that this is a result of excessive diamond wear due to the hardness of the fused silica workpiece. This smaller P-V value makes the K-value larger than the others. For few cuts and thus less diamond wear, the multiple-pellet data indicate that one might expect a surface roughness about twice that shown in row 3 (Table 2). This being the case, the K-value for fused silica would be about 5.0.

The single-pellet grinding experiments give a K-value that is valid for diamond in two grit sizes and B_4C in one size, three commonly used glasses, and several depths of cut. We believe this to be a universally applicable K-value within the precision of our measurements. We will show below that this single-pellet K-value is within the standard deviation of the multi-pellet K-value. Using the K-values in column 7,

Table 2. Results of single pellet grinding using the pneumo surfacing machine.

Row	Abrasive type	Abrasive size (μm)	Glass type	Depth of cut (μm)	σ-surface finish (μm) peak-to-valley	δ-subsurface damage depth (μm)	$K = \frac{\delta}{\sigma}$
1	Diamond	53–65	BK-7	2.0	3.0 ± 0.9	14.9 ± 2.7	5.0 ± 0.6
2	Diamond	53–65	Zerodur	2.0	2.6 ± 0.7	18.0 ± 3.7	6.9 ± 0.4
3	Diamond	53–65	Fused Silica	2.0	1.5 ± 0.7	15.1 ± 4.0	10.1 ± 2.0
4	Diamond	180–250	BK-7	2.0	4.2 ± 1.2	34.0 ± 2.4	8.1 ± 1.7
5	Diamond	180–250	Zerodur	2.0	1.2 ± 0.3	21.8 ± 2.1	18.2 ± 2.8
6	Diamond	180–250	BK-7	10.2	4.5 ± 1.3	26.2 ± 1.5	5.8 ± 1.3
7	Diamond	180–250	Zerodur	10.2	2.8 ± 0.8	Not established	N/A
8	Diamond	180–250	BK-7	30.5	8.0 ± 1.9	42.3 ± 1.3	5.3 ± 1.1
9	Diamond	180–250	BK-7	30.5[a]	7.4 ± 1.2	36.7 ± 2.0	5.0 ± 0.5
10	Boron Carbide	180–250	BK-7	10.2	5.7 ± 1.4	43.8 ± 6.0	7.7 ± 0.8
11	Boron Carbide	180–250	Zerodur	10.2	3.5 ± 0.8	26.1 ± 2.3	7.5 ± 1.0

[a] Three spark outs.

excluding the value in row 5 and taking half of row 3 as explained above, the mean value for single-pellet grinding is $K = 6.3 \pm 1.3$.

Multiple-Pellet Grinding

We investigated the mechanism of diamond grinding with multiple pellets by measuring the temporal changes in the amount of material removed, the temporal changes in surface finish, and the effect of a grinding liquid additive on the grinding properties. Our results show that for multiple pellets, the primary diamond grinding mechanism depends on the properties of the glass, the size of the diamond abrasive particles, and the mechanical properties of the bonding material holding the diamond. Izumitani[9] reached a similar conclusion from his investigation of the diamond grinding mechanism for single pellets. In addition, we have measured the depth of the damaged layer produced in fused silica for different grinding times. We used five types of diamond pellets in our investigations (Table 1). The Rockwell hardness was measured only for the metal bond; the other bonding materials contain resin.

Figure 1 shows typical multiple pellet grinding tools. Fifteen pellets, each of 6.35 mm diam, are attached to the 40.6-mm-diam stainless steel or brass tool with epoxy resin. The tool and bearing pin are hollow to allow the grinding liquid to flow through the tool onto the center part of the workpiece. The grinding liquids were pure water or water with 2 vol% Rust-Lick B additive, fed at 100 ml/min. The multiple-pellet grinding experiments were performed using a Strausbaugh[10] overhead oscillating-arm grinding-polishing machine, with workpiece on the bottom and tool on top. The arm oscillation was at 28 strokes/min with a stroke length of 16.5 mm. The spindle speed was 325 rpm, which is the upper limit of this machine. The pellet pressure was adjusted to 1.1 kg/cm^2 for each of the multiple grinding experiments.

Prior to each experiment, the grinding tools were dressed with a steel lap and a loose-abrasive slurry. The empirical rule for dressing diamond pellets is to use a dressing abrasive having a grit one or two sizes larger than the grit size of the pellet. A 30-μm aluminum oxide slurry was used to dress the 15–25-μm diamond tool. A 15-μm aluminum oxide slurry was used to dress the 6–15-μm diamond tool.

The multi-pellet diamond grinding properties were measured for Schott BK-7 and Corning No. 7940 fused silica. The Knoop hardnesses of these glasses are 595 and 630 kg/mm^2, respectively. The temporal changes in the cumulative amounts of material removed by grinding are shown in Figs. 2 and 3, both with and without the Rust Lick B additive. The grinding tool in both cases contained 20-μm diamond abrasive particles in a bronze bonding material.

Analysis of Results

Using the single-pellet data of Izumitani[9] as a guide, we believe that the no-additive data of Figs. 2 and 3 indicate a diamond-grinding mechanism that is predominantly fracture for the BK-7 material, and a combination of fracture-plastic scratching for the fused silica. At grinding times greater than 10 min, plastic scratching becomes the

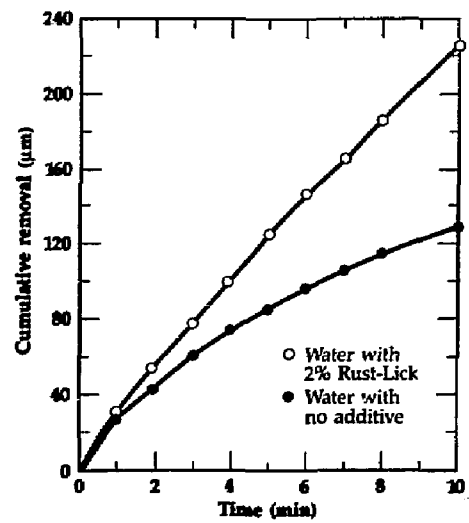


Figure 2. Cumulative material removed for BK-7 with water and water/Rust-Lick multiple-pellet grinding.

dominant grinding mechanism for the fused silica.

The experimental conditions for Izumitani's single-pellet experiment are not the same as those of our multi-pellet experiment. Differences include bond hardness, spindle speeds, loads, and diamond concentration as well as the number of diamond pellets, so no simple relation exists between the quantities of the two experiments. However, we believe that it is reasonable to make a rough comparison of the temporal trends. We measured the hardness of our bronze bonding material to be 100 on the Rockwell B scale. This corresponds to 61.5 on the Rockwell A scale, which is the hardness of several

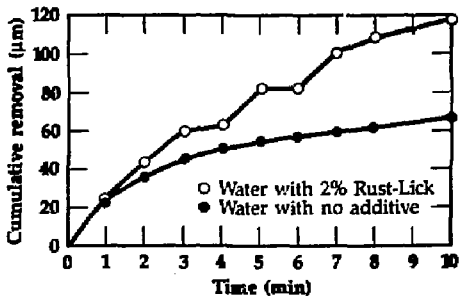


Figure 3. Cumulative material removed for Corning 9740 fused silica with water and water/Rust-Lick using multiple pellet grinding.

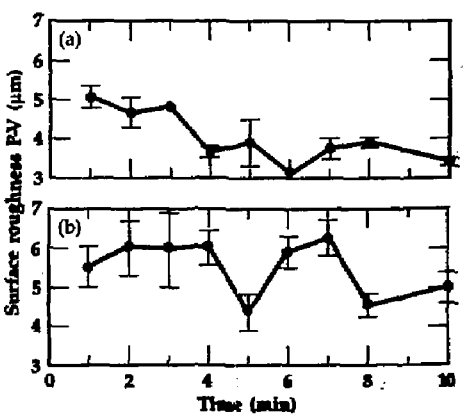


Figure 4. Surface roughness (P-V) as a function of 20-µm grinding time for BK-7; (a) with water, (b) with water/Rust-Lick.

of Izumitani's single diamond pellets.[9] Comparing the 5- and 10-min removal amounts given by Izumitani with our data, we believe that our pellets closely resemble Izumitani's No. 2 (Ref. 9, Fig. 4.38) with bonding strengths of about 30 kg/mm^2 and a degree of wear of 10.3 (Ref. 9, Table 4.3).

Additional evidence for the grinding mechanisms stated here for BK-7 and fused silica is obtained by comparing the single- and multiple-pellet temporal changes in surface finish. Our multi-pellet data (Figs. 4 and 5) show the temporal changes in surface finish for the two glasses with and without the Rust-Lick additive.

Our data show that surface finishes are approximately linear with time, although there are large error bars on the data (BK-7, with and without Rust-Lick, Fig. 4; and fused silica with Rust-Lick, Fig. 5). According to Izumitani,[9] this type of temporal behavior indicates predominantly a fracturing mechanism with minor plastic scratching. Fused silica with pure water [Fig. 5(a)] shows a slightly greater temporal dependence than the other data, indicating more plastic scratching, which was evident upon micro-

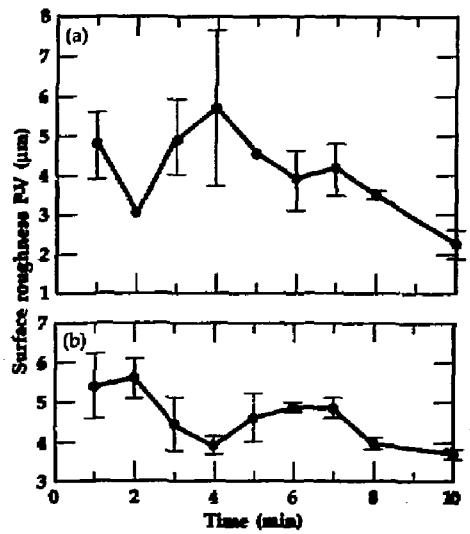


Figure 5. Surface roughness (P-V) as a function of 20-µm grinding time for fused silica; (a) with water, (b) with water/Rust-Lick.

scopic examination of the surface, especially for working times greater than 8–10 min.

Based on the discussion above and Izumitani's data, we believe that multi-pellet diamond removal for BK-7 and for fused silica with Rust-Lick is caused predominantly by a fracturing rather than by a plastic scratching mechanism. For this mechanism the diamond pellets would be continuously dressed by the ground glass waste or swarf, leading to diamond pellet wear that is linear with time. As a result the surface finish would be approximately independent of time.

As shown by data in Figs. 2 and 3, adding Rust-Lick to the polishing liquid increases the cumulative removal. The removal for the BK-7 becomes greater than that for fused silica, both with and without the additive. In normal shop practice it is more common to use a non-aqueous grinding liquid instead of water. When water is used, a hydrated layer is formed on the glass, leading to scratching instead of fracturing. A small amount of Rust-Lick retards the formation of this hydrated layer, so the predominant removal mode is fracturing. This effect is more evident for fused silica than for BK-7.

The multi-pellet grinding mechanism for fused silica with water as the grinding liquid was further examined by measuring the surface finish and damage layer depth as a function of grinding time. The same 15–25-μm cobalt/bronze-matrix diamond tool was used. The results are shown in Table 3. An additional 10 min of grinding improved the surface finish from 3.8 μm to 0.9 μm P-V, but the damage layer depth showed little decrease. We believe that this is due to the significant amounts of scratching involved in the grinding mechanism as grinding time increases. The surface will have a cosmetic, smooth finish, but little of the damaged layer is removed. For some applications a smooth surface will be sufficient. For high-power lasers and precise optical elements requiring long-term stability, our experience shows that this damaged layer is not acceptable. Time did not permit measurements similar to those shown in Table 2 for BK-7 nor measurement of the effect of Rust-Lick on fused silica.

The cumulative removal rate for fused silica depends on the dressing of the pellets. Table 4 gives the cumulative removal amounts for dressing and non-dressing conditions, each in 10-min dressing. With a dressed tool the two cumulative removals were 66.4 μm and 62.1 μm, respectively. With the same tool, but without previous dressing, the cumulative removal was only 17.3 μm. This reduced removal with time is also seen in the data of Fig. 3. We can infer that for fused silica and water as a grinding liquid the grinding mechanism is more complex than for the other cases measured. Plastic scratching appears to become an important mechanism for grinding times greater than 8 to 10 min.

Figures 6 and 7 show cumulative removal amounts for BK-7 and fused silica for several 9-μm diamond abrasives in varying resinoid bonding materials (Table 1). The linearity of the data in Figs. 6 and 7 indicates that fracture is the predominant grinding mechanism for the materials and conditions given. The surface finish data (not shown) are approximately independent of time, supporting the fracture-grinding mechanism hypothesis. Excessive diamond wear was observed for the copper-phenolic bonding material. At the completion of the fused silica measurements, the Cu-phenolic pellets were so worn that we could not perform the BK-7 measurements. Our data indicate that epoxy is ineffective as a bonding material.

Table 3. Fused silica diamond grinding properties.

Grinding time (min)	Surface finish peak-to-valley (μm)	Subsurface damage depth (μm)
10	3.8 ± 0.9	34.1 ± 1.6
20	0.9 ± 0.3	25.4 ± 2.5

Table 4. Cumulative removal for fused silica after 10-min grind.

Pellet condition	Cumulative removal (μm)
Dressed	66.4
Dressed	62.1
Not dressed	17.3

The cumulative removal amount is too small to be of practical value.

The K-value, or the ratio of the P-V surface finish to the depth of damage, can be estimated from the multi-pellet grinding experiments for fused silica. Time did not permit K-value data for the BK-7 or Zerodur. As shown in Fig. 5(a), the first few minutes of grinding, or the initial linear region, corresponds to the fracture region for fused silica and water. Taking the initial data point from Fig. 5(a) for the surface finish (4.8 ± 0.8 μm P-V), and the depth of damage from Table 3 (34.1 μm), the corresponding K-value would be 7.1 ± 0.9 for multi-pellet grinding of fused silica. This value compares favorably with the K-value of 6.3 ± 1.3 found for the single-pellet experiments (Table 2). The average of the single- and multiple-pellet K-values would yield a "universal K-value" of K = 6.7 ± 1.4. Using the data of Anderson and Frogner[11] for bound-diamond grinding, we can calculate that K = 6.1 ± 2.7 for their fused silica. Because of possible variations between their experiments and ours, this agreement is perhaps fortuitous.

Khodakov, Korovkin, and Al'tshuler[12] have reviewed the fine grinding of optical glass with a diamond tool. From their data for unspecified glasses we estimate that K = 4.3 for diamond particles between 10 to 40 μm in diameter. Their descriptions of ex-

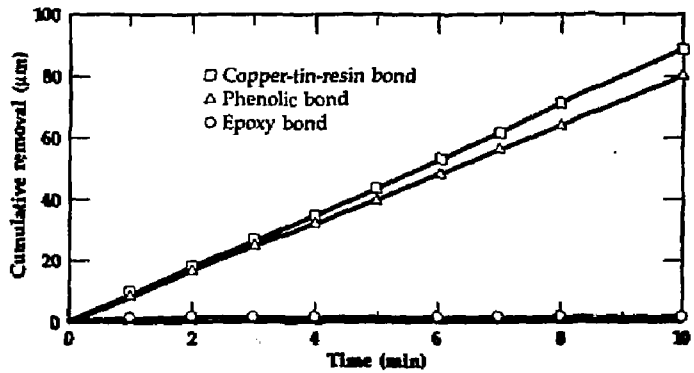


Figure 6. Cumulative material removed for BK-7 with water for 9-μm multi-pellets and various bonding materials.

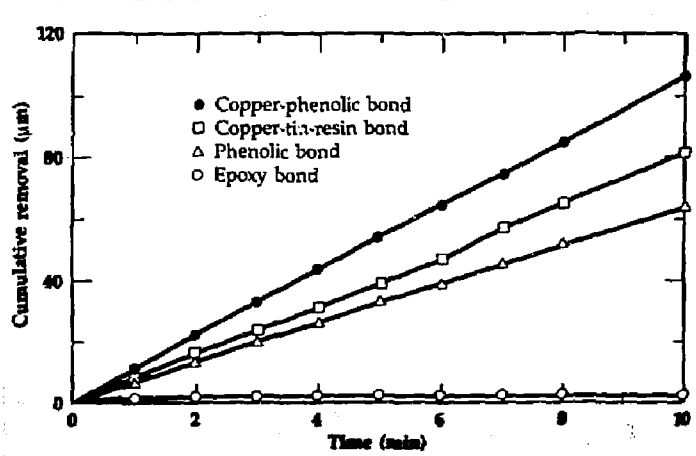


Figure 7. Cumulative material removed for fused silica with 9-μm multi-pellets and various bonding materials.

perimental details are incomplete, and their method of measuring the depth of damage is not described. Even considering these unknowns, we believe that their K = 4.3 is in reasonable agreement with our K = 6.7 ± 1.4 as reported above.

We used the taper polish technique to estimate the depth of damage in our experiments (Table 5). The polished surfaces have a characteristic pattern, with damage fissures clustering near the surface and ending at a certain depth. However, for all cases examined, a single fracture was found at a distance 1.4 ± 0.2 times deeper than the termination of the clusters. Our depth-of-damage values in this report are all based on this single, deeper fissure. If we assume that Khodakov et al. estimated their damage values relative to the shallower clusters rather than the single, deeper fracture, then their K-value would be K = 6.2 ± 0.9. This agrees much more closely with our value, but again this agreement may be fortuitous, since many details of their experiments are unknown to us.

After weighing all the evidence, we believe that our value of K = 6.7 ± 1.4 accurately describes our measurements, and is consistent with the data of Anderson and Frogner[11] as well as Khodakov et al.[12]

Conclusions

Our results show that for bound abrasive grinding the primary abrading mechanism depends on the properties of the glass and the size of the abrasive as well as on the properties of the abrasive bonding material and the grinding coolant. With hard bronze bonding material (R_B = 100) the abrading mechanism for BK-7 glass is mainly fracturing, while for fused silica it is initially fracturing, i.e., for times up to 8 min in our experiments, changing gradually to plastic scratching (viscous flow) for times greater than about 10 min. In fracturing, we find that removal rate, surface finish, and probably subsurface damage depth are independent of grinding time. Plastic scratching improves mainly the surface finish without removing subsurface damage. It appears that this improvement is caused by a wearing out and dulling of the diamond tips, a process more pronounced when grinding harder glasses such as fused silica. Adding a lubricant such as Rust-Lick B grinding fluid to the cooling water increases the duration of the fracturing mechanism by reducing friction between the cutting abrasive and the glass and thus reducing the rate at which the diamond tips wear out. Phenolic or bronze-resinoid bonding materials showed a consistent fracture-type grinding.

We have found that for fracture-type grinding there is a consistent correlation or K-value between the P-V roughness and the subsurface damage depth when using bonded-diamond or B_4C tools on fused silica, BK-7, or Zerodur. This K-value is 6.8 ± 1.7 within the accuracy of our measurements.

It seems that finer diamond grit combined with a smaller depth of cut can result in both finer surface roughness and a lesser depth of subsurface damage. This may lead to the use of several tools having successively finer sizes of grit in a deterministic automatic grinding operation. Probably the use of a single tool throughout the entire grinding operation is not optimal, especially when a large stock removal is required.

Machinery Development

A critical factor in controlling the depth of subsurface damage in bound abrasive grinding appears to be the ability of the process to proceed from brittle to ductile removal of glass. Brittle removal occurs at greater depths of cut with higher material removal rates and is characterized by considerable fracturing and attendant subsurface damage.

Table 5. Maximum damage depths for fused silica after various grinding processes (after Anderson and Frogner).[11]

Abrasive tools		Maximum pit depth (μm)	Maximum fracture depth (μm)
Grit size (μm)	Material bond		
75 ± 15	Resin	15	62
20	Metal	5	25
40	Nickel pad	6	55

Ductile removal, on the other hand, occurs at very shallow depths of cut with lower removal rates and is characterized by considerable plastic deformation and relatively less subsurface damage. Reducing the subsurface damage will allow us to reduce, or in some cases eliminate, iterative polishing steps for achieving the final figure and surface. Consequently, we are pursuing the design and construction of precision equipment, which would allow us to investigate the grinding of glass in a ductile fashion.

We are beginning to explore two different approaches to modeling the transition between brittle and ductile grinding. The first approach, based on fracture mechanics, postulates that the transition from brittle to ductile behavior is governed by force, or, more specifically, by a stress. The second approach, stemming from the chemistry of polishing, suggests that the transition is governed by the depth of cut; because of hydration, the properties of the glass surface are different from the bulk properties, leading to a different behavior in the surface layer. It is not yet clear which process offers the best hope of modeling the cutting process during ductile grinding.

A survey of research in glass grinding[12-15] has led us to the conclusion that controlling the performance of the grinding head is critical to the maintenance of ductile glass removal. Our plan for FY 86 therefore called for a grinding head designed and built specifically for ductile grinding. We would then be able to experimentally characterize the conditions under which ductile grinding occurs. To this end we have obtained and are making appropriate modifications to a spindle and motor drive suitable for ductile grinding of glass. This unit was originally designed and built for the PERL machine.[16]

Certain requirements must be met by a grinding head intended for ductile grinding of a brittle material such as glass. For example, the grinding wheel must spin with minimum runout, either normal or tangential, at the point of contact with the workpiece. Also, the spinning wheel must not excite vibrations in the structure of the machine. Both requirements are a consequence of the need to grind a workpiece to an accurate form at cutting depths close to, but not exceeding, the transition from ductile to brittle removal. Good balance, high stiffness, and a high-quality axis of rotation are characteristics important to any grinding head, but are particularly relevant when grinding brittle materials.

Imbalance has a number of deleterious effects during grinding. First, it causes runout of the periphery of the wheel, which varies as a function of speed. Unless compensation is applied, varying the speed of the wheel can cause an error in the depth of cut or the size of the workpiece. Imbalance can also cause the wheel to cut at one spot on its periphery instead of distributing wear uniformly over the entire wheel. Finally, imbalance causes an excitation force and concomitant vibration within the structure of the machine tool. This in turn can cause a degradation in the finish of the workpiece being ground.

Machine stiffness is important for many reasons. High stiffness reduces the shifting of the axis of rotation, and hence the runout of the wheel periphery due to any residual imbalance. It also reduces the deflections of the wheel with respect to the workpiece resulting from the small fluctuating forces that accompany grinding. Further, high stiffness is known to reduce the problems associated with chatter, a self-excited vibration induced between the workpiece and the tool.

A high-quality axis of rotation is required of a grinding head. A spindle that produces no error motion as a result of imbalance may still exhibit error motion due to other influences, which might include geometric imperfections in the spindle components, dynamic effects from the spindle air bearings, and the influence of the drive. In general these effects are thought to be small relative to imbalance effects.[17]

Description of the Spindle

The grinding head we have obtained is particularly suited for grinding brittle materials (see Figs. 8 and 9). An air-bearing spindle is used for the actual axis of rotation. This spindle has radial and axial stiffness of 1.5 and 0.5 μin. per pound of applied force,

respectively. The drive for the spindle is a brushless dc motor, whose rotor is cantilevered directly from the spindle rotor. In this design there are no bearings in the motor to degrade the axis of rotation defined by the air bearing.

Because of its brushless design most of the heat developed in the motor is generated in the stator. A temperature-controlled water jacket surrounds the stator to remove this heat before it migrates into the grinding head structure. If the generated heat were to enter the grinding head structure, it would cause dimensional change, which would, for example, introduce uncertainty in the depth of cut. The motor is driven with a sinusoidal voltage instead of the usual square wave to reduce degradation of the axis of rotation.

Initial Evaluation of the Spindle

We performed an initial evaluation of the spindle to assess its applicability to the grinding of glass. Our greatest concern was whether we could balance the grinding head well enough to preserve the high quality of the axis of rotation inherent in the air-bearing spindle. Most balancing is performed by placing a transducer on a stationary part of the grinding head and measuring acceleration or velocity resulting from imbalance. However, our goal is to reduce the excursions of the rotor to a level difficult to sense by an accelerometer.

We therefore chose to measure runout, which is the relative change in the gap between the rotor and stator as the rotor spins (Fig. 10), by using a capacitance gauge that measures displacement between the rotor and the stator. Such a measurement is complicated by factors other than imbalance that

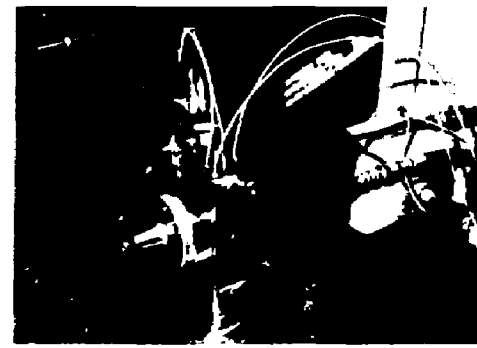
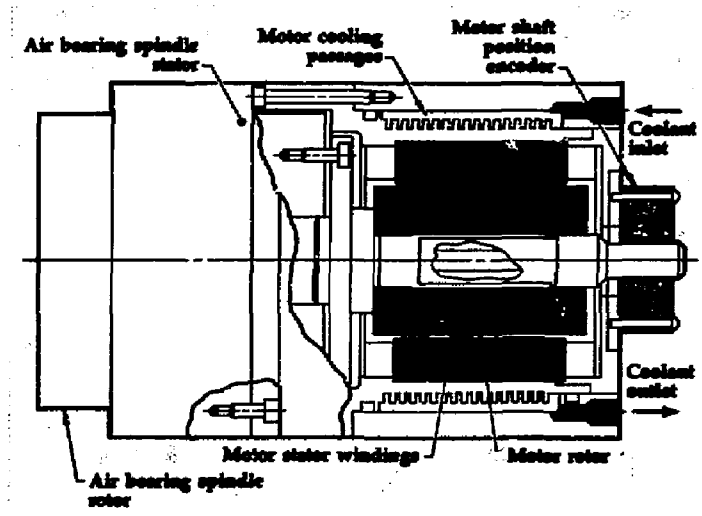


Figure 8. Grinding head with air-bearing spindle.

Figure 9. Schematic cross section of air bearing spindle with integral motor.



may change the rotor-stator distance. For example, if the rotor is not perfectly round, then the gap will vary as the rotor spins even if the rotor is perfectly balanced. The scheme employed here is to make measurements at more than one speed in order to separate those effects that do not vary with speed, such as rotor roundness, from those that do, such as imbalance. This could be seen as taking a measurement at zero spindle speed and then subtracting this signal from that obtained at grinding speed. For a high quality air bearing the difference between the two signals is due to imbalance.

In our tests we first rotated the well balanced spindle at different speeds and measured the runout. Next we placed an imbalance mass of 7 g on the spindle at a radius of 38 mm and repeated the runout measurements. The results (Fig. 11) show runout in μin. vs spindle speed in rpm. The dashed curve shows the data for the case with no added imbalance. Note that both curves start with about the same amount of runout at low speeds. We feel that in the low speed range we are observing error motion not due to imbalance. Therefore we can graphically subtract this error motion from our data to leave only the component due to imbalance.

From this initial evaluation we think that the spindle can meet the demands of ductile grinding of glass. The runout can be limited to 1 or 2 μin. and the exciting forces to a value resulting in a comparable structural displacement.

Future Work

We are in the process of visiting various centers where research on ductile grinding is under way. Our purpose is to define the parameters that determine whether glass grinding will occur in a brittle or a ductile fashion. We feel that we have a good understanding of the required hardware, but we require further knowledge concerning such factors as the techniques of dressing the grinding wheel, the fluids used during grinding, and the specific process parameters.

We believe that a fundamental understanding of the transition between ductile and brittle material removal is crucial to controlling that transition in order to achieve economy and quality in manufactured optical elements. This understanding will develop from well-founded analytical studies corroborated by experiments.

To this end we will review the application of classic fracture mechanics to propagating surface cracks and then progress to more in-

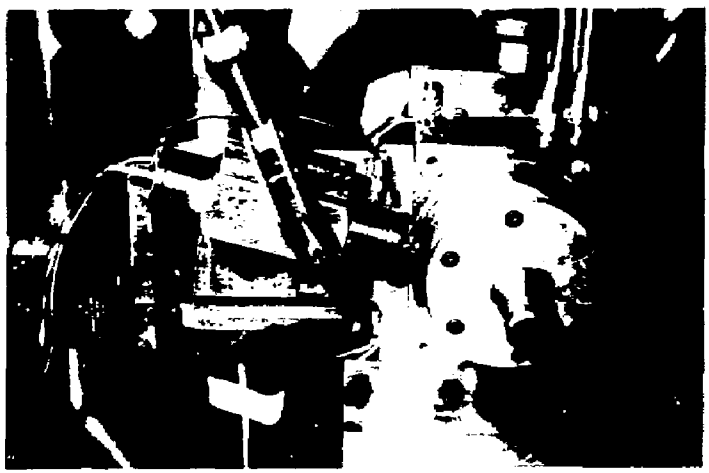


Figure 10. Capacitive gauge in place to measure runout.

tricate modeling to achieve a better grasp of the brittle-ductile transition. We will also strive to explain and model the transition between brittle and ductile failure based on the hypothesis that the surface of the glass workpiece has undergone hydration from reactions with water-based coolants. Because of these reactions, the surface properties differ from those of the bulk glass.

Improved grinding machinery resulting from the FY 86 work will enhance our ability to perform the carefully controlled experiments required for this research. The new spindle and grinding head assembly, mounted in place on the Laboratory's recently completed CBN numerically controlled machine tool, offer a good foundation on which to proceed. This installation accommodates precision experiments that are conveniently sized. We anticipate substantially lower research costs with this system as compared to the previously proposed designs built around much larger precision tools. Eventually we will expand our capability to a larger scale. However, based on our results to date, numerous experimental iterations lie ahead as we search for optimal grinding parameters.

Outside Contacts

We have exchanged information on the single- and multiple-pellet grinding mechanisms with T. Izumitani of Hoya Corporation, Tokyo, Japan.

Also, we have exchanged information on ductile grinding with the following experts and institutions:

- M. Miyashita and F. Hashimoto, with whom we held discussions at the 2nd International Grinding Conference concerning LLNL's attempt to develop the capability of grinding glass in a ductile mode.
- T. Nakagawa, who during a visit to LLNL gave a presentation on the development of a cast-iron bonded-diamond grinding wheel.
- North Carolina State University, which has recently initiated a program to investigate ductile grinding of brittle materials. We are collaborating with researchers there, having recently sent them a sample of glass with a low coefficient of expansion on which to perform grinding. We will subsequently analyze the ground sample here at LLNL for subsurface damage.

Financial Status

Engineering research funds spent on this project during FY 86 total $250 000.

Acknowledgments

We would like to acknowledge the contributions of N. J. Brown, J. B. Bryan, D. F. Edwards, H. J. Hansen, and R. S. Stolcis to this report.

1. P. P. Hed, N. J. Brown, D. F. Edwards, and R. S. Stolcis, *Fine Grinding of Optical Glass with Bound Abrasives*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985).
2. J. B. Bryan and D. L. Carter, *Precision Contouring of Brittle Materials*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985).
3. Schott Optical Glass Inc., Dureya, PA; and Corning Glass Works, Corning, NY.
4. P. P. Hed and D. F. Edwards, *Preventing Rust Accumulation on Iron Optical Lapping Tools*, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-95406 (1986).

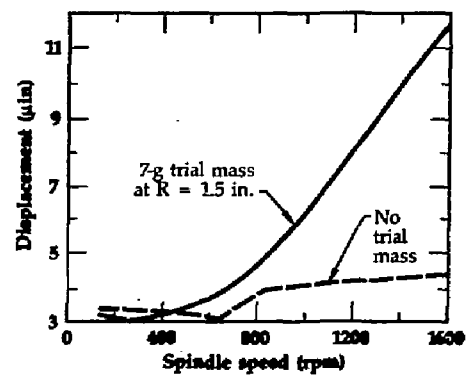


Figure 11. Spindle imbalance test data, showing net runout displacement.

5. Model MSE-326 (Pneumo Precision Inc., Keene, NH).
6. Rust-Lick B-CD-2 water-soluble corrosion inhibitor coolant concentrate (Devcon Corp., Drexel Hill, PA).
7. Frater's Facsimilike, Drexel Hill, PA.
8. Surfanalyzer Model 2000 (Federal Products Corp., Providence, RI).
9. T. S. Izumitani, *Optical Glass* (Kyoritsu Shuppan Co., Ltd., Publishers, Tokyo, Japan, 1984).
10. Model GUR2 (R. Howard Strasbaugh, Huntington Beach, CA).
11. D. S. Anderson and M. E. Frogner, *A Method for the Evaluation of Subsurface Damage*, presented at the Optical Fabrication and Testing Workshop, Cherry Hill, NJ, June 12–13, 1985.
12. G. S. Khodakov, V. G. Korovkin, and V. M. Al'tshuler, "Physical Principles of the Fine Grinding of Optical Glass With a Diamond Tool," *Sov. J. Opt. Technol* 47 (1986).
13. J. Yoshioka et al., "Ultraprecision Grinding Technology for Brittle Materials," ASME Paper PED V16, 207 (1985).
14. M. Miyashita et al., "New Concept of Grinding Technology for Predictability of Manufacturing Operation," SME Paper MR86-645 (1986).
15. A. B. Van Groenou and J.D.B. Veldkamp, "Grinding Brittle Materials," *Phillips Tech. Rev.* 38, 105 (1978–79).
16. R. R. Donaldson and D. C. Thompson, "Design and Performance of a Small Precision CNC Turning Machine," *Ann CIRP* 35, 373 (1986).
17. *Axes of Rotation. Methods for Specifying and Testing*, ANSI/ASME Standard B89.3.4M-1985 (1985).

Study of Zero CTE of Graphite/Epoxy Composites

W. W. Feng
Materials Test and Evaluation Section
Engineering Sciences Division

D. A. Schauer
Weapons Engineering Division

The design of composites with very low coefficients of thermal expansion (CTE) is discussed. The magnitude of the CTE of a composite laminate is usually coupled with its mechanical properties. In this report the thermomechanical properties of multidirectional laminates, consisting of plies with arbitrary orientations and stacking sequences, are obtained with the laminate theory. Tests were performed to determine the thermomechanical properties of unidirectional composite laminates and to verify the analytical results for a variety of composite laminates. These tests measured the elastic moduli, Poisson's ratios, and CTEs. A differential dilatometer was used to provide the required accuracies for measuring low CTEs. All test results agree with the analytic i predictions. Finally, the authors propose future investigations on other subjects tha. affect the design of very low CTE composites.

Introduction

The large variety of properties of fiber composites requires methods of analytical evaluation to enable the designer to choose suitable composite materials. All such analyses are based on models that simplify reality so as to make mathematical analysis possible. In this report we develop such a mathematical model to determine the coefficient of thermal expansion (CTE) of a fiber composite laminate. We have two objectives. The first is to apply the laminate theory to determine the ply orientations for a desired CTE. The second is to develop accurate test methods to measure the thermomechanical properties of a composite laminate and then verify the analytical results. We used the laminate theory to obtain the thermomechanical properties for multidirectional laminates consisting of plies with arbitrary orientations and stacking sequences. Experiments were then performed to verify the validity of the laminate theory and the numerical results. The analytical results provide the longitudinal and transverse moduli of elasticity, the longitudinal and transverse CTEs, and Poisson's ratios for a $[\theta_4/-\theta_4]_s$ laminate. All analytical predictions agree with the experimental results; the comparisons are presented.

In order to apply the theory with facility, we developed a computer program. When designing composite laminates in the future, one can use the computer program to study the many choices of stacking sequences and ply orientations. Numerical studies offer in a very short time many design possibilities that corresponding experimental study would take years to produce.

This project has direct applica' on to the Strategic Defense Initiative Prog m where a stiff, lightweight, zero (or very ow) CTE platform is required. Researchers in other LLNL projects, such as for the earth penetrator weapon, Sabot, and free electron laser, will also find that the theoretical analyses, computer programs, test methods, and data obtained from this project are useful. Besides these programs, many LLNL projects have become heavily involved in using new materials, especially fiber composites. This project will benefit them as well. The results and the method obtained from this project will also benefit the composites community, since a major application of composites is space structures where thermal expansion and thermal fatigue are critically important.

The other important areas that affect the thermomechanical properties of a composite's structure are the residual stress during fabrication, the failure criterion, the

maximum allowable temperature change, size effect, and zero CTE fabrication techniques. These are listed as possible future research areas.

Technical Status

Analysis

The theoretical work here is a study of the effects of ply orientations and stacking sequences on CTEs and on the mechanical properties of the composite. The laminate theory to be developed in this section was employed. The constitutive relationship for composite materials is

$$\boldsymbol{\sigma} = \mathbf{Q}\,\boldsymbol{\epsilon} \quad , \tag{1}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ are the stress and strain tensors of order 2, respectively; $\mathbf{Q}$ is the stiffness tensor of order 4. For a single-ply composite where the plane stress assumption is used, the explicit forms of these quantities, written in terms of their components, are

$$\boldsymbol{\sigma}^T = [\sigma_{11}\ \sigma_{22}\ \sigma_{12}] \quad , \quad \boldsymbol{\epsilon}^T = [\epsilon_{11}\ \epsilon_{22}\ \epsilon_{12}]$$

$$\mathbf{Q} = \begin{bmatrix} \dfrac{E_{11}}{1-\nu_{12}\nu_{21}} & \dfrac{\nu_{21}E_{11}}{1-\nu_{12}\nu_{21}} & 0 \\ \dfrac{\nu_{12}E_{22}}{1-\nu_{12}\nu_{21}} & \dfrac{E_{22}}{1-\nu_{12}\nu_{21}} & 0 \\ 0 & 0 & \mu_{12} \end{bmatrix} \quad , \tag{2}$$

where E is the modulus of elasticity; μ is the shear modulus of elasticity, and ν is Poisson's ratio. The subscripts denote directions. The ply-axis, denoted by 1- and 2-axes, are shown in Fig. 1. The 1-axis, in the longitudinal direction, coincides with the fiber direction. The 2-axis, in the transverse direction, is perpendicular to the fiber direction.

When both mechanical and thermal loads are applied to the single-ply laminate, $\boldsymbol{\epsilon}$ can be decomposed into two parts:

$$\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^T + \boldsymbol{\epsilon}^M \quad , \tag{3}$$

where $\boldsymbol{\epsilon}^T$ represent the strains due to thermal loads, and $\boldsymbol{\epsilon}^M$ represent the strains due to mechanical loads. The thermal strains are related to the CTEs ($\boldsymbol{\alpha}$) and to the temperature changes (ΔT) by

$$\boldsymbol{\epsilon}^T = \boldsymbol{\alpha}\,\Delta T \quad . \tag{4,}$$

For a single-ply composite laminate, $\boldsymbol{\alpha}$ can be written in terms of its components:

$$\boldsymbol{\alpha} = [\alpha_{11}\ \alpha_{22}\ \alpha_{12}] \quad . \tag{5}$$

The material properties $\boldsymbol{\alpha}$ and $\mathbf{Q}$ are in the ply axis. A rotated-axes system, denoted by 1'- and 2'-axes, is also shown in Fig. 1. The angle of rotation is ϕ. These quantities in a rotated-axes system are related to the quantities in the ply-axes system by

$$\begin{aligned} \boldsymbol{\alpha}' &= T^t\,\boldsymbol{\alpha}\,T \\ \mathbf{Q}' &= \mathcal{T}^t\,\mathbf{Q}\,\mathcal{T} \quad , \end{aligned} \tag{6}$$

where T and $\mathcal{T}$ are the second- and fourth-order tensor transformation matrices, respectively.

The mechanical properties of a laminate with n plies of various orientations and stacking sequences can now be obtained. In this case, the thermal strain value $\boldsymbol{\epsilon}^T = 0$ is assumed. The stress results ($\mathbf{N}^L$) and the in-plane strains for a laminate are related by

$$\mathbf{N}^L = \mathbf{A}\,\boldsymbol{\epsilon}^L \quad , \tag{7}$$

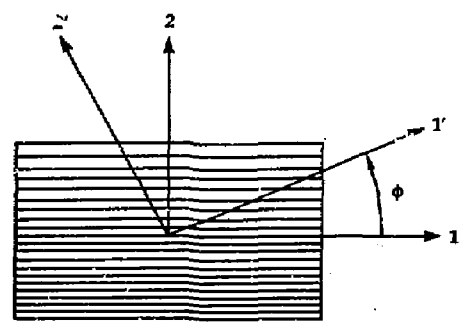


Figure 1. Definition of the ply-axis system and the rotated-axes system.

where the stress resultants are defined by

$$\mathbf{N}^L = \int_h \boldsymbol{\sigma}\, dh \quad , \tag{8}$$

and the in-plane moduli of a multidirectional laminate (A) are defined by

$$A = \sum_{k=1}^{n} \mathbf{Q}'_k\, h_k \quad , \tag{9}$$

where h_k is the thickness for the kth ply.

Equation (7) can be rewritten as

$$\boldsymbol{\epsilon}^L = \mathbf{a}\, \mathbf{N}^L \quad , \tag{10}$$

where $\mathbf{a}$, the in-plane compliance matrix of a multidirectional laminate, is the inverse of $\mathbf{A}$.

If a multidirectional laminate consists of plies with uniform ply thicknesses (h_0), then the apparent mechanical properties of the laminate are

$$E^L_{11} = 1/n\, h_0\, a_{11} \quad , \quad E^L_{22} = 1/n\, h_0\, a_{22} \quad ,$$
$$\mu^L_{12} = 1/n\, h_0\, a_{33} \quad , \quad \nu^L_{12} = -a_{21}/a_{11} \quad . \tag{11}$$

The thermal properties of a laminate with n plies of various orientations and stacking sequences can also be obtained. In this case, the assumption that the stress resultants are zero is implied, or

$$\mathbf{N}^L = \int_h \mathbf{Q}\, \boldsymbol{\epsilon}^M\, dh = 0 \quad , \tag{12}$$

and

$$\int_h \mathbf{Q}\, (\boldsymbol{\epsilon}^L - \boldsymbol{\epsilon}^T)\, dh = 0 \quad . \tag{13}$$

With Eqs. (1), (7), and (9), Eq. (13) reduces to

$$\sum_{k=1}^{n} \mathbf{Q}'_k\, \boldsymbol{\alpha}^L\, h_k\, \Delta T - \sum_{k=1}^{n} \mathbf{Q}'_k\, \boldsymbol{\alpha}'_k\, h_k\, \Delta T = 0 \quad , \tag{14}$$

or

$$\boldsymbol{\alpha}^L = \frac{\sum_{k=1}^{n} \mathbf{Q}'_k\, \boldsymbol{\alpha}^L_k\, h_k}{\sum_{k=1}^{n} \mathbf{Q}'_k\, h_k} \quad . \tag{15}$$

Equation (15) indicates that the thermal properties of a composite laminate are functions of both mechanical and thermal properties of the single-ply unidirectional laminate.

Computer Program

Using the formulation described in the theoretical study, the researchers developed an interactive computer program. The program solves various aspects of mechanics problems in composite material structures, including the thermomechanical properties. The values of the material properties for a unidirectional ply are assigned at the beginning of the program. The program asks for other necessary input values. Some numerical results are presented.

Experimental Work

Five panels of T-300/F263 fiber composite laminates with $[\theta_4/-\theta_4]_s$ orientation and stacking sequence, as shown in Fig. 2, were prepared. Five θ values, 0°, 10°, 20°, 30°, and 45°, were chosen. Two mechanical and two thermal specimens, each with their longitudinal axes coinciding with the 1-axis, were cut from the panels. The other two mechanical and two thermal specimens with the longitudinal axes coinciding with the 2-axis were cut from the panels. Therefore, for the test specimens, θ has nine values: 0°, 10°, 20°, 30°, 45°, 60°, 70°, 80°, and 90°. The mechanical properties of a unidirectional composite laminate can be obtained from uniaxial tensile tests:

$$E_{11} = \left.\frac{\sigma_{11}}{\epsilon_{11}}\right|_{\theta = 0^\circ} \quad , \quad E_{22} = \left.\frac{\sigma_{11}}{\epsilon_{11}}\right|_{\theta = 90^\circ} \quad ,$$
$$\mu_{12} = \left.\frac{\sigma_{11}}{2(\epsilon_{11} - \epsilon_{22})}\right|_{\theta = 45^\circ} \quad , \tag{16}$$
$$\nu_{12} = -\left.\frac{\epsilon_{22}}{\epsilon_{11}}\right|_{\theta = 0^\circ} \quad .$$

The CTEs of a unidirectional composite laminate can also be obtained from one-dimensional thermal expansion tests:

$$\alpha_{11} = \frac{\epsilon_{11}}{\Delta T} \quad \theta = 0^\circ \quad ,$$

$$\alpha_{22} = \frac{\epsilon_{11}}{\Delta T} \quad \theta = 90^\circ \quad , \tag{17}$$

$$\alpha_{12} = 0 \quad .$$

The thermomechanical properties for a unidirectional composite are determined with the $[0_8]_s$, $[90_8]_s$, and $[45_3/-45_4]_s$ specimens. The results of the tests for T-300/F263 composite laminates are

$$\begin{aligned} E_{11} &= 21.1 \times 10^6 \text{ psi} \quad , \\ E_{22} &= 1.26 \times 10^6 \text{ psi} \quad , \\ \mu_{12} &= 0.65 \times 10^6 \text{ psi} \quad , \\ \nu_{12} &= 0.246 \quad , \\ \alpha_{11} &= -0.25 \times 10^{-6}/^\circ\text{C} \quad , \\ \alpha_{22} &= 34.1 \times 10^{-6}/^\circ\text{C} \quad . \end{aligned} \tag{18}$$

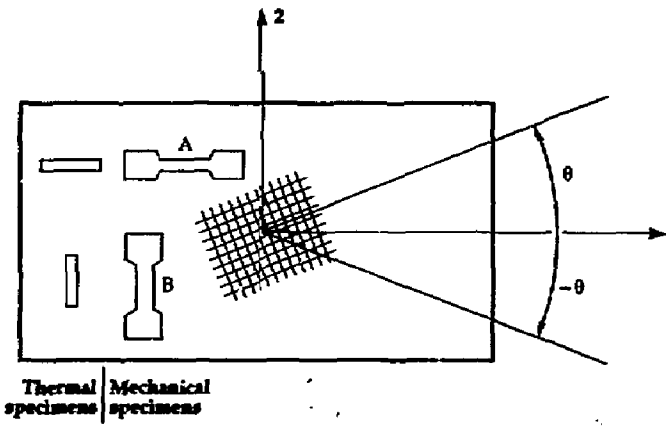


Figure 2. Diagram of the thermal and mechanical test specimens.

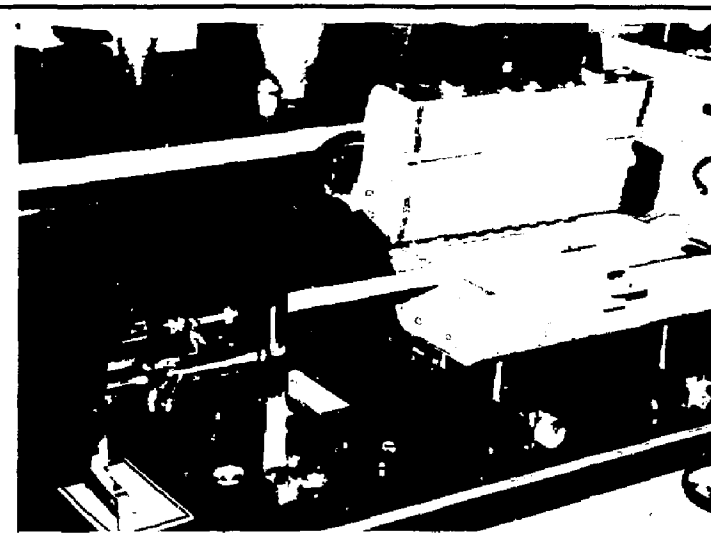


Figure 3. The differential dilatometer for measuring low CTE.

The other specimens were also tested; their results were used mainly to verify the theoretical predictions. The theoretical predictions and the experimental results are compared in the next section.

From the preceding formulation, only one-dimensional thermomechanical tests are needed to determine the fundamental material properties of a unidirectional composite laminate. Some very small quantities are measured in these tests. Accurate tests and data acquisition systems were required and were developed for this project. In the uniaxial tensile tests, the results are accurate to 0.1%. In order to provide the required accuracies for measuring low CTEs, a differential dilatometer is used to simultaneously measure the differential expansion of two specimens. A picture of the apparatus is shown in Fig. 3. The differential dilatometer was checked by using calibrated specimens. The calibration indicates a precision of 0.03 to 0.06 × 10^{-6} per °C for the coefficient of thermal expansion. To test thermal expansion of composites more accurately, we compared the CTEs of a specimen with that of the standard quartz specimen, while the linear variable differential transformer (LVDT) chamber temperature was maintained at 40°C.

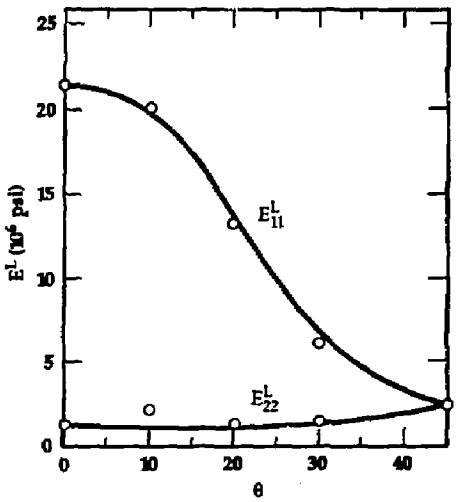


Figure 4. Theoretical and experimental elastic moduli for $[\theta_4/-\theta_4]_s$ laminates.

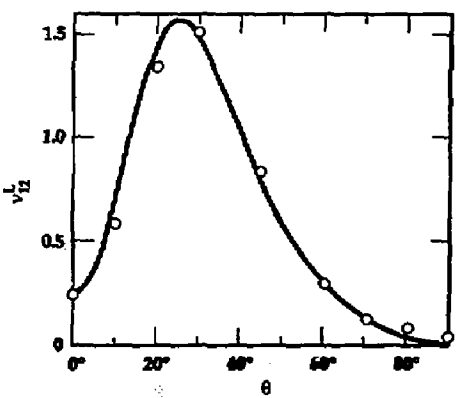


Figure 5. Theoretical and experimental Poisson's ratios for $[\theta_4/-\theta_4]_s$ laminates.

Results

The moduli of elasticity and Poisson's ratios for $[\theta_4/-\theta_4]_s$ graphite/epoxy composite laminates are shown in Figs. 4 and 5, respectively. The CTEs for $[\theta_4/-\theta_4]_s$ graphite/epoxy composite laminates are shown in Fig. 6. The lines indicate the

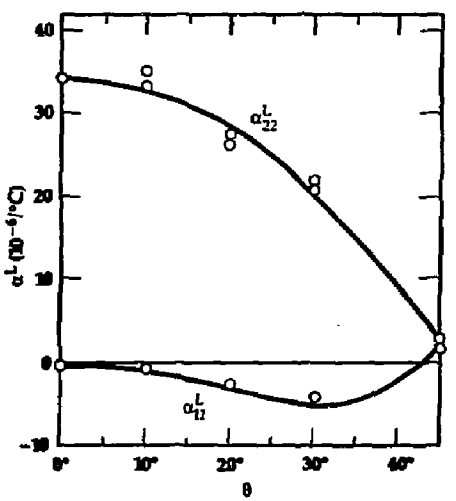


Figure 6. Theoretical and experimental CTEs for $[\theta_4/-\theta_4]_s$ laminates.

analytical results, and the data points indicate the test results. From these curves, the analytical predictions agree well with the experimental results. The results also indicate that a very low CTE graphite/epoxy fiber composite laminate can be obtained.

Future Work

The very low CTE for graphite/epoxy composites has been obtained theoretically and the results verified experimentally. However, during the course of study, the investigators identified the following subjects that affect the design of very low CTE composites:

- Residual stress for epoxy-resin composites. A micromechanics model for studying the effect of curing on the residual thermal stress of epoxy-resin composites will be developed. The thermorheological constitutive equations for epoxy and the effect of residual strain on CTE will be determined.
- Failure criterion. The failure surface for composite materials will be determined. In addition, a new criterion based upon the strain invariants will be developed. Experiments will be performed to determine the failure criterion quantitatively.
- Maximum allowable temperature change. When the residual stress and failure surface of a composite is determined, the maximum allowable temperature change for a composite before fracture will be determined precisely. Experiments will be performed to verify the analytical results. The nondestructive evaluation methods, such as acoustic emission, will be used to determine the onset of failure.
- Size effect. A simple flat panel, cut into small test samples for mechanical and CTE evaluation, is relatively easy to fabricate. However, a larger structure is usually fabricated from several panels with several right-angle joints. This more complex structure may not exhibit the same thermal expansion behavior as the small specimens. The large-scale CTE apparatus will be used to determine the CTE of a finished large part approximately 30-in. long.
- Zero CTE fabrication techniques. Several other factors can also be important in achieving a zero CTE composite. Epoxy volume fraction, high modulus fibers, low CTE matrix fillers, and precise alignment of each laminate ply all can affect the overall composite CTE. Further effort in each of these areas will determine the practical limit in achieving a very low CTE.

Each of the foregoing subjects will be a research topic by itself; each topic will be investigated in the future. As far as all goals stated in the original proposal of this project and the current applications of low CTE to the Strategic Defense Initiative Program are concerned, this project is complete.

Outside Contacts

We have constantly interacted with the personnel in the chemistry and mechanical departments that are developing new materials for R Program.

A paper entitled "The Design of Low CTE of Graphite/Epoxy Composites" has been accepted for publication by the Society of Manufacturing Engineers, Composites in Manufacturing 6, Anaheim, CA, January 19–22, 1987.

Financial Status

The total expenditure for this project in FY 86 was $304 000.

High-Temperature Metal Alloy Radiant Property Measurements in Conjunction with Advanced Surface Spectroscopy

M. Havstad
Energy Systems Engineering Division

W. McLean
Laser Isotope Separation Program

The purpose of this work is to study the radiative and optical properties of pure liquid metal surfaces using both a state-of-the-art radiation property measurement system and the recently developed techniques of surface analysis. These techniques allow us to analyze in detail the atomic composition of a metal surface. Research reported to date has not utilized these tools, so sample materials have been impure and of unknown surface composition. In FY 86, we designed and fabricated an apparatus, which, for the first time, will allow complete radiative property measurements and surface spectroscopy to be done in the same device. Our system employs argon ion sputtering, Auger electron spectroscopy (AES), and ultra high vacuum techniques and makes radiative property measurements as a function of angle, wavelength, and temperature. After assembly and shakedown, we used the apparatus to make two sets of measurements. The data are being analyzed. We have also initiated theoretical work to compare our data to predictions derived from the Fresnel equations and free electron theory. In FY 87 we intend to make measurements on several additional materials of interest in many technologies and develop a predictive capability by working with existing theoretical relations.

Introduction

The thermal radiative properties of liquid metals are important in a variety of technologies being studied both within and outside the Laboratory. At LLNL the Laser Isotope Separation (LIS) process, under development as the nation's future uranium enrichment capability, has the greatest need for improved understanding of molten metal thermal radiation. In proposed LIS systems, where molten metal films must flow continuously, local temperatures must be maintained above the freezing temperature of the metal. The liquid metal radiant properties—the fraction of blackbody radiation emitted (emissivity), and the fraction of incident radiation reflected (reflectivity)—determine local temperatures and, thus, whether or not film flow is maintained and system design is acceptable, or whether freezing occurs and system redesign or modifications are required.

Lawrence Livermore National Laboratory laser and electron-beam welding projects, where radiative heat loss influences performance, will also potentially benefit from studies of liquid metal radiant properties. Outside the Laboratory some other areas of note are vacuum arc remelting, arc welding, and designs of advanced nuclear reactor cores where, as with the LIS process, radiative properties combine with other factors to determine energy flows, temperature distributions, and, ultimately, device performance.

Radiative properties of materials have been measured or estimated with a variety of techniques since at least the end of the Stone Age, when man began to work with metals and blacksmiths sought to know the temperatures of their work pieces. Measurement techniques advanced as the need for accuracy increased or as more difficult environments were encountered. By the mid 1960s, the standard references on this subject[1–3] all discussed the dependence of

radiative properties on surface state, composition, structure, roughness, and other factors. However, discussions of surface condition were limited to crude descriptions of surface polishing or roughening techniques and bulk material elemental assays. These characterizations have long been inadequate and have been the source of large uncertanties and expensive margins in process designs. Furthermore, with liquid metals, the surface material's reactivity, and, in some cases, its gettering ability make surface specifications commonly found in the literature irrelevant. Molten metals form smooth surfaces, so the critical specification is surface composition. The strong gettering abilities of uranium and stainless steel, in particular, require that the environment in which surface property measurements are made be very tightly controlled. Visual observation of hot and especially molten metals confirms that radiative properties vary widely with surface composition. Figure 1 shows the reported data for the emittance of uranium. The data, which extend nearly to the melting temperature at 1406 K, vary widely because each citation is for a different but unknown surface composition. Likewise, Fig. 2 shows reported emittance data for stainless steel. Once again, variable and undetermined surface compositions account for a range of reported values, which renders the whole of the data nearly meaningless.

The thrust of this research project has been to bring quantitative surface analysis tools to radiative property measurement so that meaningful data on known surfaces are obtained. These tools are argon ion sputtering, which purifies metal surfaces, and Auger electron spectroscopy (AES), which precisely determines surface composition. With these techniques, we may obtain and report the radiative properties of pure materials of known composition.

Technical Status

Experimental Work

In response to the inadequacies cited above in both the reported data and the tech-

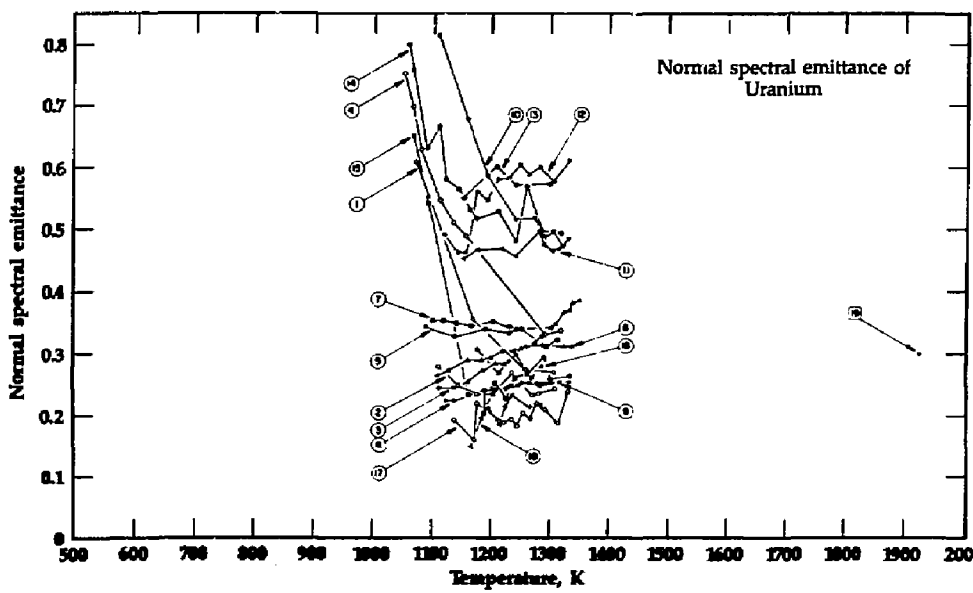


Figure 1. State of the data—uranium (from Touloukian and DeWitt, *Thermophysical Properties of Matter*, vol. 7, IFI/Plenum, New York, 1970, p. 835. Used with permission.).

niques heretofore applied, we designed a molten metal radiative property system with surface analysis capability. Such a design had not been attempted until now for a number of reasons. Quantitative surface analysis was not developed until about 1969 and was not commercially available until about 1972. Even now, surface analysis equipment is not widely available—perhaps 300 Auger spectrometers are in use in the U.S.—and it remains expensive ($100 000–250 000). Furthermore, work with molten metals and surface equipment required a host of other expensive and complex equipment, including, for example, ultra-high vacuum (UHV) pumping capability, ion sputtering guns, and all-metal vacuum systems.

To control surface composition, our only option was to place the radiative property measurement system in the same vacuum vessel as the surface analysis equipment. Figure 3 shows a plan view of the side-by-side layout. On one side of the vacuum vessel we have a surface analysis plane and on the other a radiative property measurement plane. We translate molten samples, always in vacuum (approximately 1×10^{-9} Torr or better), between the two planes to do a complete set of measurements. In the surface analysis plane we sputter clean the molten sample surface until the material surface is pure. Sputter cleaning of metals is particularly important because bulk impurities tend to be lighter than the bulk material and therefore float to the surface as the metal melts. Even with bulk material as clean as 30 ppm, the lighter impurities congregate into an optically thick surface scum on melting, which must be removed to obtain worthwhile measurements. The large port on the right-hand side of the surface analysis plane holds the Auger electron spectrometer. The ion gun focuses on the center of the plane and is arrayed near the optimal

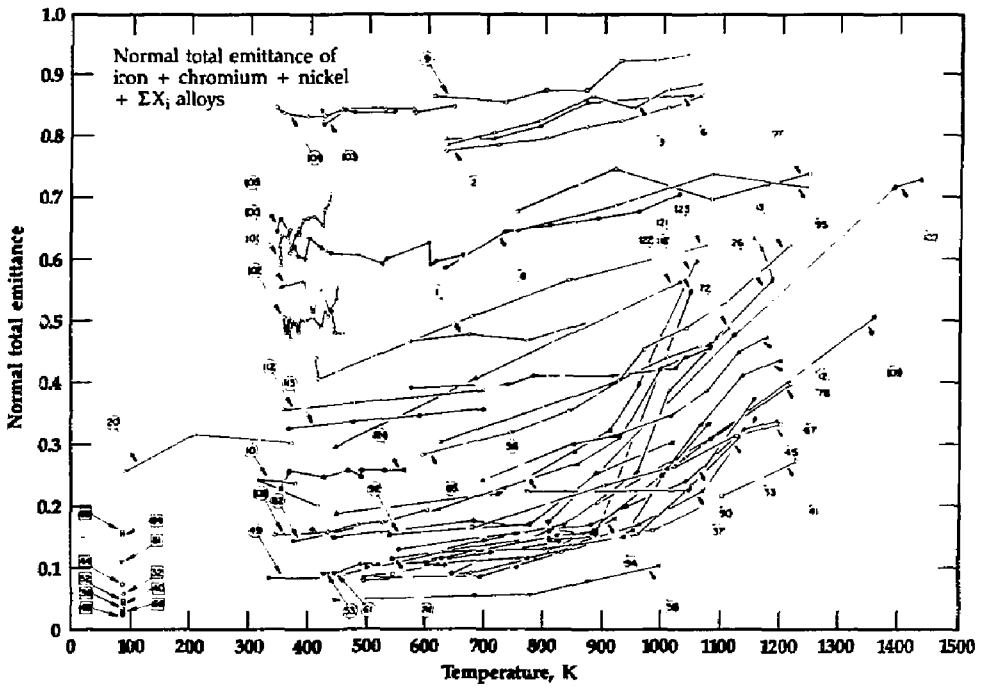


Figure 2. State of the data—stainless steel (from Touloukian and DeWitt, *Thermophysical Properties of Matter*, vol. 7, IFI/Plenum, New York, 1970, p. 1220. Used with permission.).

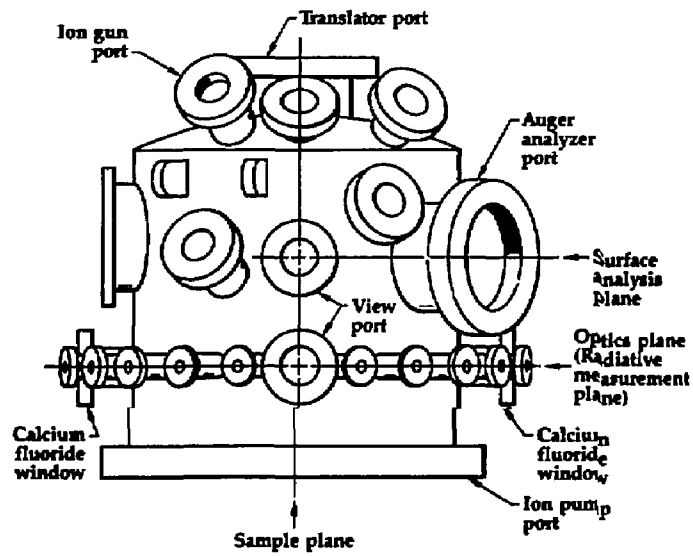


Figure 3. Side-by-side layout of the radiative property measurement system and the surface analysis equipment. (View looking down on chamber and liquid metal surface plane.)

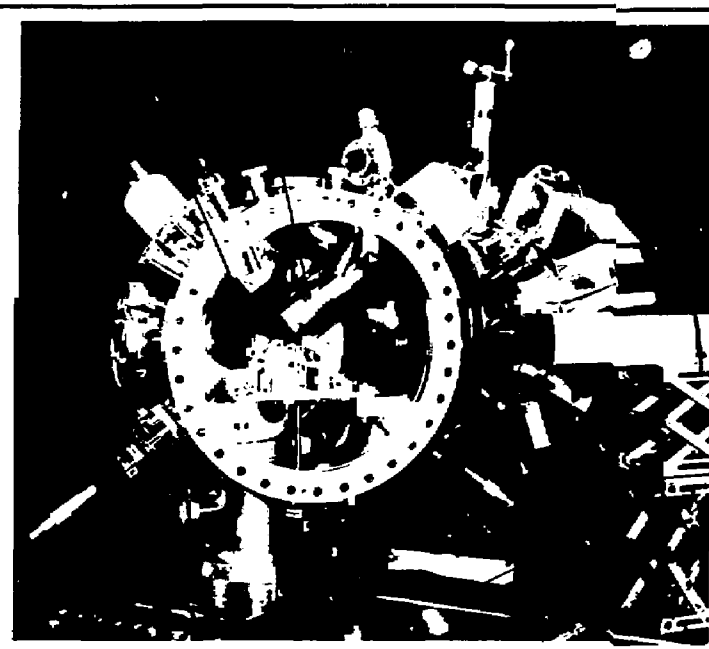


Figure 4. Optics of the radiation property measurement system.

sputter yield angle so that surfaces can be cleaned with the least effort.

Figure 4 shows the vacuum chamber with the surface analysis equipment in place. The larger cylinder extending almost into the middle of the chamber is the Auger electron analyzer. The smaller cylinder is the ion gun. The vacuum chamber is mounted to both the optical table and the large gate valve shown in the bottom of the photograph.

Figure 5 shows the vacuum chamber with the optical system in place. The aluminum cylinders, arrayed in an arc over the sample, hold the mirrors at 15° intervals; two axis tilt manipulators allow the mirrors to be aligned from outside the vacuum. Also shown in Fig. 5 are several of the various optical paths from which measurements are obtained. The mirrors on either side of the center of the chamber, where the gold reference mirror is in place, are pivoting and attached to vacuum feedthroughs to allow mirror rotation from outside vacuum.

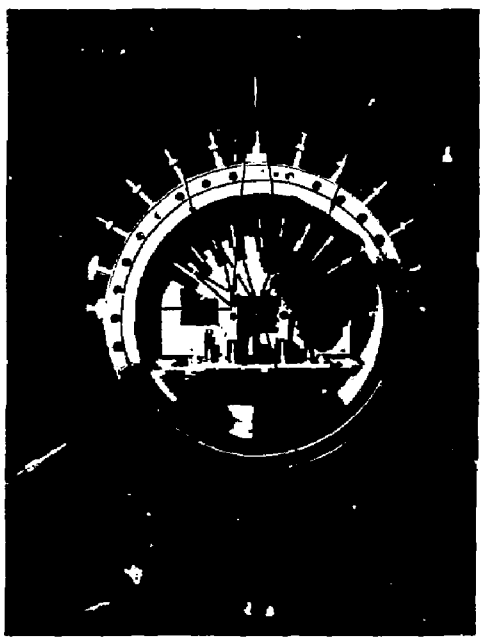


Figure 5. Vacuum chamber with surface analysis equipment in place.

Figure 6 shows the specifics of the radiation property measurement plane. The 10 gold mirrors above the molten sample allow radiant property measurements to be made as a function of angle in 15° increments from 15° to 75° off normal. To measure the fraction of the incident light reflected by a surface, the light source at the left of the figure is imaged at the sample surface with the mirrors as shown. The reflected light is then collected and imaged on the detector on the right side of the apparatus. A gold reference mirror is substituted for the sample periodically to allow calculation of sample reflectivity. The detector converts the optical signal incident on its face to an electronic signal, which is processed by a lock-in amplifier and read digitally. For measurements where the detector output is very nearly linear we obtain sample reflectivity from:

$$\rho_{sample}(\theta,\lambda,T) = \rho_{gold}(\theta,\lambda,300\ \mathrm{K}) \times \frac{E_{sample}(\theta,\lambda,T)}{E_{gold}(\theta,\lambda,300\ \mathrm{K})}, \quad (1)$$

where λ signifies a wavelength dependence, θ an angular dependence, and T a temperature dependence. E_{sample} and E_{gold} are the detector output voltages, after processing, due to the light reflected off the sample and the gold reference mirror, respectively. Only the sample shows a temperature dependence because the gold mirror is held at ambient temperature throughout all measurements.

For emissivity measurements, the sample's emitted light is imaged on the detector, as is the output of the blackbody by a separate set of mirrors. The gold coated chopper blade causes the detector to see alternately the sample and the blackbody. In this way we compare sample emission to blackbody emission and may then calculate sample emissivity. Calculating emissivity from the data is only slightly different than calculating reflectivity. However, a small correction must be made in the emissivity experiment to account for the fact that the optical path from the blackbody to the detector is not identical to the optical path from

the sample. The two optical paths are identical in the reflectivity experiment. In both cases, bandpass filters in front of the detector allow measurement of particular wavelengths between 0.4 μm and 10 μm. The system uses reflective optics to avoid the complications caused by the dependence of index of refraction on wavelength and calcium fluoride windows to allow 0.4- to 10.0-μm radiation to enter and leave the vacuum chamber.

Measurement of both reflectivity and emissivity are redundant and therefore allow a check on measurement accuracy. From conservation of energy arguments, we can show that

$$\rho = 1 - \epsilon \quad ,$$

where ρ is reflectivity and ϵ is emissivity. Thus the amount of discrepancy between our reflection results and one minus the emission results gives a direct indication of our experimental error.

Our apparatus yields radiant property data as a function of angle, wavelength, and temperature. To obtain more commonly reported values such as total hemispherical reflectance (a reflectivity measurement for light at all wavelengths from all angles) we simply perform the relevant integration of our angular and/or specular data to obtain the desired result. The raw data set is more complete and contains a wealth of information on radiative properties, optical constants, and material structure, but the integrated quantities are used more widely.

To summarize the experimental status at this writing we have built and used the apparatus to make reflectivity measurements on a polished solid copper sample. These measurements are being reduced and the copper data are being compared to measurements made on the same sample by a standard

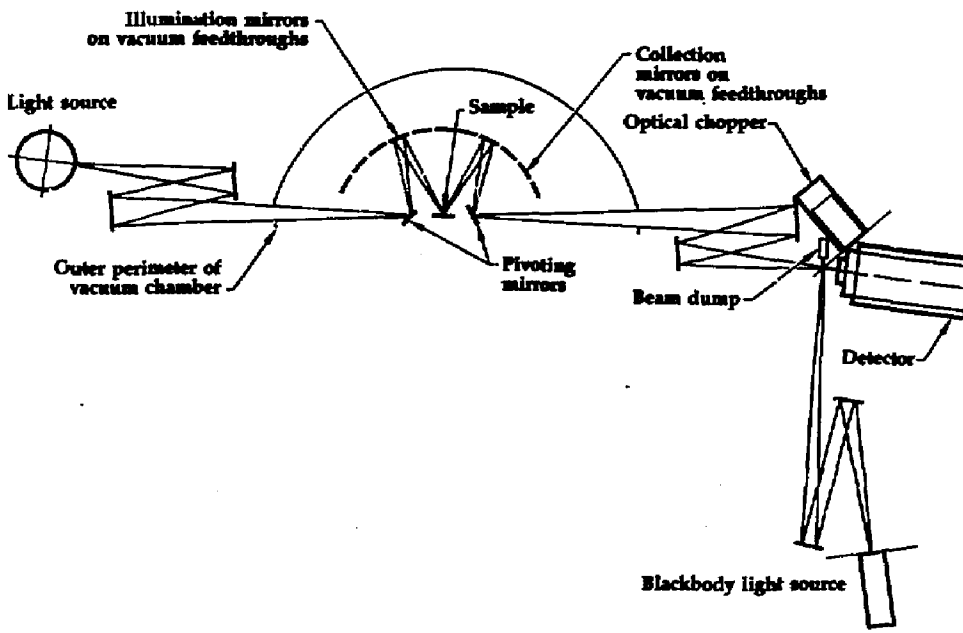


Figure 6. Vacuum chamber with optical system in place.

gonioreflectometer, which operates in air at room temperature. Once we have verified that our apparatus checks out using a dirty but otherwise identical sample, we can move on to measurements with pure materials.

We began this fiscal year with a rough concept on the back of an envelope and we completed it with a functional apparatus that finally brings qualitative surface analysis to radiation property measurement.

Theoretical Work

While our experimental efforts yield reflectivity and emissivity values of use in studies of a number of emerging, advanced technologies, our apparatus can also be used to determine the fundamental optical constants of liquid metals. By performing our reflection measurements with several key considerations in mind we will be able to compute the complex index of refraction of the liquid metals we study and open up a large number of possibilities for theoretical work.

The Fresnel equations relate the complex index of refraction to the reflectivity of plane-polarized and perpendicularly polarized light as a function of the angle of incidence. These equations are given below:

$$\rho_{\perp} = \frac{a^2 + b^2 - 2a\cos\theta + \cos^2\theta}{a^2 + b^2 + 2a\cos\theta + \cos^2\theta} \quad ,$$

$$\rho_{\parallel} = \rho_{\perp} \times \frac{a + b - 2a\sin\theta\tan\theta + \sin^2\theta\tan^2\theta}{a^2 + b^2 + 2a\sin\theta\tan\theta + \sin^2\theta\tan^2\theta} \quad , \qquad (2)$$

where

$$2a^2 = \sqrt{(n^2 - k^2 - \sin^2\theta)^2 + 4n^2k^2} + (n^2 - k^2 - \sin^2\theta) \quad ,$$

$$2b^2 = \sqrt{(n^2 - k^2 - \sin^2\theta)^2 + 4n^2k^2} - (n^2 - k^2 - \sin^2\theta) \quad .$$

The complex index of refraction consists of n, the refraction index, and k, the extinction coefficient. The refraction index accounts for the bending or refraction of light as it strikes an interface and k, the extinction coefficient, accounts for the damping of electromagnetic waves as they pass through conductors. By working with polarized light in our reflection apparatus or by using one of several other considerations, we may deduce the optical constants, n and k. We are presently reviewing the several techniques that allow computation of n and k and will select the method most suited to our apparatus and accuracy needs. In the most promising method we simply add a polarizing filter to the optical system in front of the detector.

Once we have computed the optical constants we may begin to develop theories that help us understand optical behavior in more detail. For example, we can use free electron theory to relate the optical constants to the free carrier density and collision frequency as follows:

$$n^2 - k^2 = 1 - 2\frac{Ne^2}{2\pi m^*}\frac{1}{\nu^2 + \gamma^2}$$

$$nk = \frac{\gamma Ne^2}{\nu 2\pi m^*}\frac{1}{\nu^2 + \gamma^2} \quad , \qquad (3)$$

where

ν = frequency of light,
γ = electron collision frequency,
n = index of refraction,
k = extinction coefficient,
e = electronic charge,
m^* = effective electron mass,
N = free electron density.

One goal of this work is to obtain and use the two basic free electron parameters of pure metals to predict the properties of binary alloys. This is a formidable task but if completed would represent a significant advance in our understanding. A successful predictive model would also reduce the data gathering required to provide infor-

mation necessary for studies of emerging technologies.

In summary, our theoretical work continues in parallel with the experimental efforts. We are reviewing the available techniques for obtaining the complex index of refraction as well as the expected range of validity of the free electron model and perturbations to it.

Future Work

For the coming fiscal year, we plan to use the apparatus to make radiant property measurements on several materials of wide interest and then to expand on the theory of liquid metal radiant properties. Finally, we will compare predictions obtained from the theory to our data.

At present, some possible materials for radiant property measurements are aluminum, stainless steel, and silver. Weld modeling efforts at LLNL and numerous other research sites provide motivations for the aluminum and stainless-steel work. The silver data may be of use to those investigating fabrication processes that occur during solid-state bonding. In addition, we propose to do measurements of a binary alloy over its full range of possible compositions. If successful, the results may be applied where binary materials (or polyatomic alloys, where all but two materials are present in only minor concentrations) are used in welding or fusing.

The existing theoretical work for pure materials has been applied with varying levels of success to both solid and molten materials of approximate surface composition. A comparison using our new pure material data will be illuminating. Furthermore, an extension of theoretical work to binary alloys appears to be straightforward.

Outside Contacts

We have enjoyed a close working arrangement with Dr. S. Self of the Department of Mechanical Engineering at Stanford University. A presentation on our progress at Stanford is planned for the Fall of 1986 and we will continue to receive feedback from Stanford on new applications and refinements to the apparatus.

We were contacted by Precision Cast Parts of Portland, Oregon, to provide them with radiative properties for materials used in jet engine vane parts. They have been unable to obtain data for their materials from other sources.

Financial Status

This project was jointly funded by the Engineering Research Program and the LIS Program. We spent about $75 000 ($45 000 for supplies and expenses, and $30 000 for manpower) of Engineering Research Program funds. Additional funding of $100 000 was supplied by the LIS Program.

1. R. Siegel and J. R. Howell, *Thermal Radiation Heat Transfer* (McGraw-Hill Book Company, New York, NY 1972).
2. E. M. Sparrow and R. D. Cess, *Radiation Heat Transfer* (McGraw-Hill Book Company, New York, NY 1966).
3. V. S. Touloukian and D. P. DeWitt, *Thermophysical Properties of Matter*, vol. 7 (IFI/Plenum, New York, NY 1970).

NMR Imaging for Materials Characterization

P. E. Harben
Engineering Research Division
Electronics Engineering Department

A. Maddux
Nuclear Energy Systems Division
Electronics Engineering Department

R. L. Ward and W. Boyle
Chemistry Department

Nuclear magnetic resonance (NMR) imaging tools have proven unique and powerful in medicine for imaging many regions of the body. Very little of this technology has been used in nonmedical disciplines, however. Our objective was to develop and use NMR imaging techniques to characterize the properties of materials of particular interest to researchers in Laboratory programs.

We modified a solids spectrometer built at Sandia National Laboratories to perform NMR imaging. We also built two sample probes and added such features as quadrature phase detection and spin-echo formation to enhance our imaging capabilities. Several test samples containing mobile water, including small water flasks and peapod sections, were imaged with good results. We also imaged with some success water in such materials as epoxy composites and welded tuff rock cores. However, water in these samples sometimes became bound to the material walls (thus rendering the water immobile), and we have so far only been able to image samples containing some mobile water. With additional research, however, we anticipate that solids can be successfully imaged with NMR technology.

Introduction

Nuclear magnetic resonance (NMR) imaging has become a powerful medical research and diagnostic technique. Although relatively new,[1] NMR scanners have proven superior to computed axial tomography (CAT) scanners in non-invasively imaging many regions of the body. Furthermore, most medical imaging devices use ionizing radiation to probe a patient's body, thus often exposing the patient to significant radiation doses. Since the NMR signal is in the radio-frequency band, it contains no harmful ionizing radiation.

The unique information contained in a spatially encoded NMR signal (i.e., one in which the location of nuclei giving off a signal can be identified) has opened several exciting research areas in medicine, ranging from early cancer detection to *in vivo* studies of the metabolic process. The usefulness of the NMR signal for imaging depends on two independent relaxation times (the T_1 and T_2 relaxation times), the nuclear density of the material being imaged, and the method used to spatially encode the information. By suitably designing an NMR imaging experiment, one can cause the received signal to be more or less dependent on these quantities. This capability is very important because nuclear relaxation times are strongly influenced by the chemical environment; hence, one can "set" the experiment to be strongly dependent on the chemical structure. Indeed, observations of the widely differing T_2 relaxation times between normal and cancerous tissue,[2] in conjunction with the discovery of a method to spatially encode the NMR signal, first led to the current boom in medical NMR research.

As unique and powerful as NMR imaging tools have become in the medical community, the use of NMR technology in examining materials has scarcely been

explored?[3-6] At present, most medical NMR research has concentrated on imaging hydrogen nuclides in mobile water or fats. Indeed, hydrogen nuclei give a very strong NMR signal and are in humans and animals in large concentrations. A large number of materials have large hydrogen nuclei concentrations as well as nuclei not present in biological systems. Not all nuclei exhibit the NMR phenomenon or exhibit it to a very measurable degree, but, as shown in Table 1, a large number do. The potential of NMR imaging in non-medical areas is enormous.

Our goal this fiscal year was to explore some nonmedical applications of NMR imaging. We completed converting a 4.7-T NMR solids spectrometer located at Sandia National Laboratories to NMR imaging and imaged test samples of general interest and material samples of programmatic interest. This report details our accomplishments. We first discuss the NMR phenomenon and the conversion of the spectrometer to NMR imaging; we then present some results of imaging experiments.

Technical Status

NMR Imaging Background

The NMR phenomenon, a quantum mechanical concept, occurs only in nuclei with an odd number of protons or neutrons. These nuclei can be thought of classically as containing a separation of charge and therefore act as tiny dipoles. These dipoles have a magnetic moment, and, when placed in an external magnetic field, tend to align with the externally applied field, creating net bulk magnetization in the sample. When a second magnetic field perpendicular to the first is applied, the bulk magnetization of all the nuclei tips away from the direction of the first field. The application of the second field excites the nuclei; that is, they are moved from a low to a high energy state. For this to happen, the secondary field must be applied at a specific frequency, which, for a given static field strength, is unique for each nucleus. The nuclei in the sample precess about the new external field (i.e., the combined static and secondary fields) at the frequency of the secondary field and, as the nuclei relax to their lower energy states, give rise to a measurable electromagnetic signal. This is the NMR phenomenon.

The specific frequency of precession for the nuclei is directly proportional to the external magnetic field strength

$$\omega = \gamma B_0 \quad ,$$

where ω is the resonant frequency of precession or Larmor frequency, γ is a constant of proportionality that is unique for each nucleus, and B_0 is the external field strength. When an rf pulse is applied at the resonant frequency, the nuclei begin to precess about the combined applied field direction at the resonant frequency. The rf pulse duration, if chosen properly, can cause the net applied field direction to be rotated 90° to the static, external field. (Such a pulse is called a 90° pulse.) After the rf pulse is turned off, the nuclei realign themselves with the static, externally applied field. The return to the original alignment, termed the relaxation, involves two processes: the spin-spin relaxation (caused by the interaction of neighboring nuclei to dephase the bulk magnetization), designated T_2, and the spin-lattice relaxation (caused by the growth of magnetization parallel to the static magnetic field), designated T_1. As mentioned earlier, these relaxation times depend on material composition.

When another magnetic field, called a field gradient, is imposed on the static magnetic field such that the total magnetic field

Table 1. Some nuclei exhibiting the NMR phenomenon.

Strongest NMR Response	Strong NMR Response	Moderate NMR Response				Weak NMR Response
^{1}H	^{7}Li	^{23}Na	^{113}Cd	^{51}V	^{125}Te	^{13}C
^{3}He	^{11}B	^{63}Cu	^{151}Eu	^{71}Ga	^{195}Pt	^{29}Si
^{19}F	^{31}P	^{93}Nb	^{45}Sc	^{121}Sb	^{59}Co	^{77}Sc
^{203}Tl	^{87}Rb	^{141}Pr	^{69}Ga	^{187}Re	^{81}Br	^{127}I
^{205}Tl	^{117}Sn	^{27}Al	^{113}In	^{57}Mn	^{129}Xe	
	^{119}Sn	^{65}Cu	^{185}Re	^{79}Br	^{207}Pb	

strength B_o changes linearly with distance, the Larmor frequency also changes because the Larmor frequency is proportional to B_o. This concept is the key to spatially encoding the NMR signal. Figure 1(a) shows how a gradient field applied in the x direction gives rise to a changing Larmor frequency across a specimen containing two samples of water, thereby locating the positions of the two water samples along the x axis. If the experiment is repeated many times, where each time the axis along which the gradient is

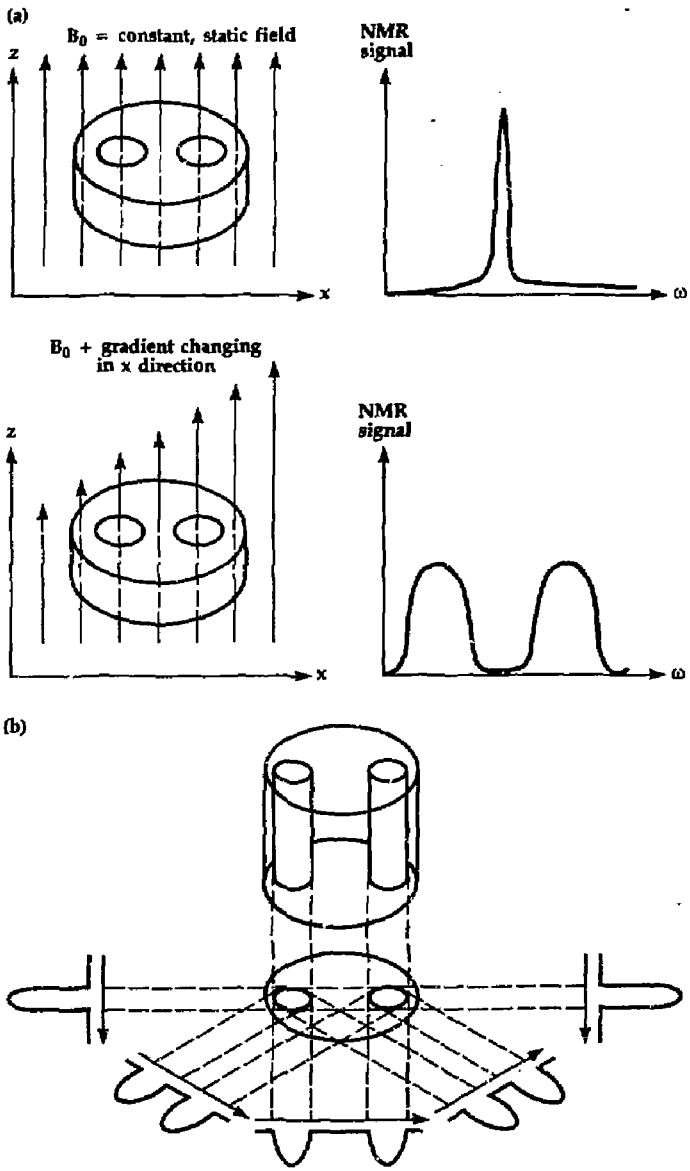


Figure 1. Application of field gradients to produce spatially encoded NMR signals. (a) The static external field (B_o) in the z direction, when Fourier transformed, produces a single NMR response at the Larmor frequency. Note that the two water samples (represented by the circles) are not differentiated. However, imposing a gradient field that changes linearly with respect to distance causes the Larmor frequency of nuclei in the sample to change in response to the changing field strengths. (This change in field strength is represented by the difference in heights of the arrows.) Thus, the Fourier-transformed NMR signal is a one-dimensional projection with two distinct components corresponding to the location of the samples along the x axis. (b) By rotating the axis along which the gradient is applied, a series of one-dimensional angular projections, each showing the location of the two water samples, is obtained. These projections can then be processed to compute a two-dimensional image.

applied [Fig. 1(b)] is rotated, the resulting collection of one-dimensional projections can be processed, using algorithms similar to those used in CAT scanners, to recover a complete, two-dimensional distribution of the NMR signal.

Three-dimensional distributions are obtained by selectively rf-exciting thin slices of a sample and then applying the technique discussed above to recover the two-dimensional distribution of the NMR signal within that slice. This method of spatially encoding and image-reconstructing the NMR signal—the so-called projection reconstruction technique—represents only one of several ways to accomplish NMR imaging. All methods do, however, involve the application of field gradients along three mutually orthogonal axes.

Figure 2. The Nalorac 4.7-T vertical bore superconducting magnet. This magnet produces the static magnetic field essential for NMR characterization.

NMR Imaging Conversion

The high-resolution solids spectrometer made available to us uses a Nicolet 1280 computer and pulse programmer. The heart of the spectrometer is the Nalorac 4.7-T vertical bore superconducting magnet shown in Fig. 2. We modified the spectrometer to produce two-dimensional NMR images by contracting with Nalorac to build *x*-, *y*-, and *z*-axes gradient coils. We then built the gradient driver interface to the computer and pulse programmer (Fig. 3). The software was

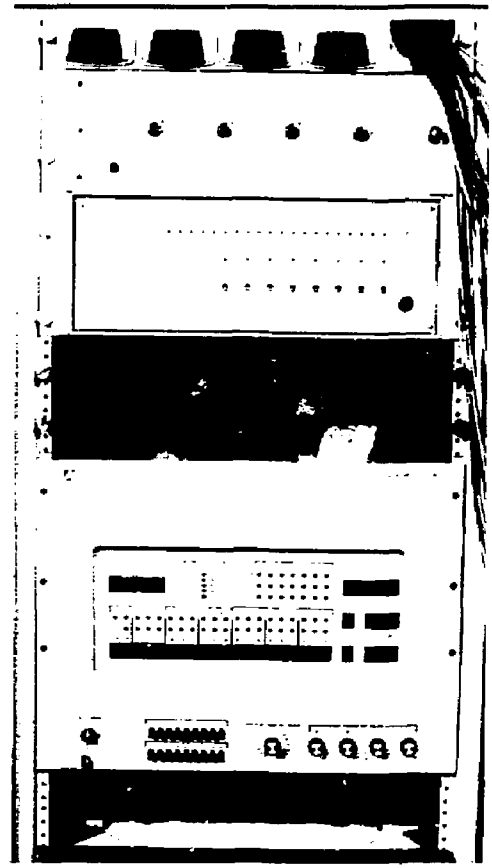


Figure 3. The computer and gradient driver interface. The interface permits the computer and software to drive the *x*, *y*, and *z* gradient coils.

modified to control the gradient drivers, thereby allowing computer control of the gradient fields during NMR experiments. A two-dimensional Fourier transform algorithm already available on the computer was used to reconstruct images from the raw data.

Early tests with our "bare bones" imaging capability showed some shortcomings. Poor signal-to-noise (S/N) ratios, spectral folding, and a non-flat power spectrum led us to implement a quadrature phase detection technique. Thus, we could increase the S/N ratio of our pulses by a factor of 1.4, eliminate folding, and flatten the power spectrum.

The key improvement in our NMR imaging capability was the implementation of a spin-echo pulse sequence during data acquisition, which also significantly improves our S/N ratio. The pulse sequence is shown in Fig. 4. The 90° ($\pi/2$) rf pulse [Fig. 4(a)] produces an NMR response [Fig. 4(d)] even before x and y field gradients [Figs. 4(b) and 4(c)] are applied. This fact, coupled with long rise and fall times in the 90° pulse due to interfering eddy currents created in magnet core metal (something most manufacturers now avoid), prevents the experimenter from gathering data until after a significant amount of the NMR signal has already passed. The exponential decay of the NMR signal results in significant losses and, therefore, a poor S/N ratio. The introduction of a second rf pulse [the 180° or π pulse of Fig. 4(e)] causes a realignment of magnetic moments in the nuclei to occur, thus forming an "echo" of the original signal well after the second rf pulse is concluded. Because the echo has already been spatially encoded (i.e., the x and y gradient pulses have been applied), the experimenter does not have to wait for eddy currents to decay and can acquire data during the full echo period [Fig. 4(f)], thus dramatically increasing the S/N ratio.

NMR Imaging Results

We confined our investigations to imaging the hydrogen nuclei of mobile water in solids. This provided the easiest starting point for our efforts and also allowed us to take advantage of the larger body of existing research information in the medical community.

One of the first NMR images we obtained was of three 75-µl glass spheres filled with

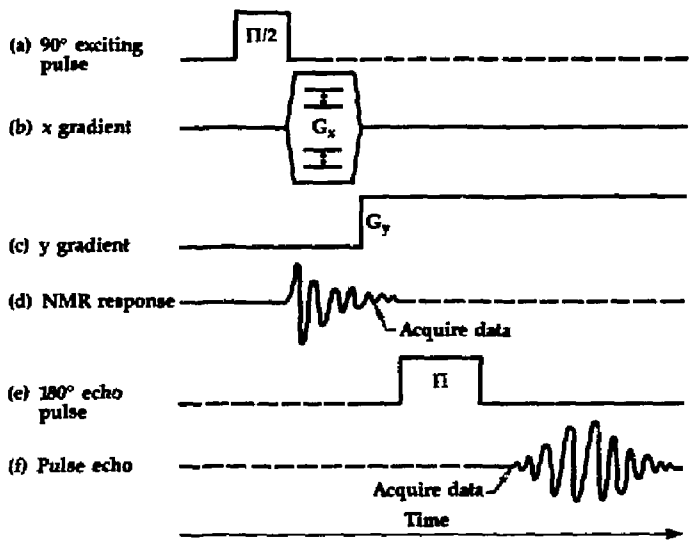


Figure 4. The spin-echo pulse sequence. The $\pi/2$ pulse at (a) stimulates an NMR signal (d). However, this response begins before the x gradient (b) and y gradient (c) fields are applied. Therefore, a π pulse is applied (e), which causes an "echo" of the original NMR signal to be created (f). However, since the gradients have already been applied, the experimenter can collect data as soon as the NMR response is detected.

water [Fig. 5(a)]. The image we obtained is shown in Fig. 5(b) without gradient fields applied and in Fig. 5(c) with gradient fields applied. Without gradients the NMR signal is not spatially encoded; hence, the signal resonates about a narrow range of Larmor frequencies (due to inhomogeneities in the main magnetic field). When two independent gradient fields are applied, the spatially encoded NMR signal produces the contour images seen in Fig. 5(c). We attribute the slight distortions of the circular shape to nonlinearities inherent in the main magnetic field and to errors in calibration of the gradient field strengths.

Another early test consisted of imaging a 3-mm-diam glass flask filled with water at half maximum gradient field strength [Fig. 6(a)] and maximum gradient field strength [Fig. 6(b)]. Although the flask shape is very distorted at half strength, at full gradient field strength an accurate image of the water within the flask is produced. In all imaging applications, the gradient field strength must be maximized to produce an accurate image.

To test the system's ability to image a more complex mobile water distribution we imaged a section of a peapod, including its stem [Fig. 7(a)]. Figure 7(b) shows the final image we obtained. The break in the contours in the stem region corresponded to an actual break in the peapod stem. Image detail was good and the S/N ratio was high, indicating a significant presence of mobile water.

We obtained epoxy composite samples from the Nondestructive Evaluation Group of the Mechanical Engineering Department. The first group of samples consisted of epoxy cylinders cured with a small fraction of injected water, which formed minute bubbles within the final samples. These samples were found to contain no signifi-

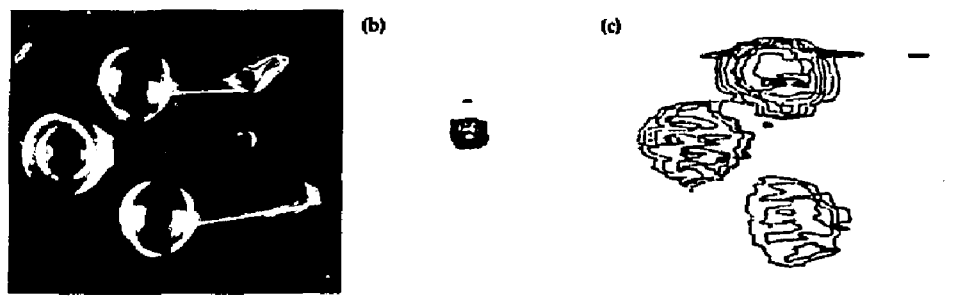


Figure 5. Use of gradient fields to provide an image. (a) Spheres used in the imaging experiment. (b) Image without gradient fields applied. (c) Image with gradient fields applied.

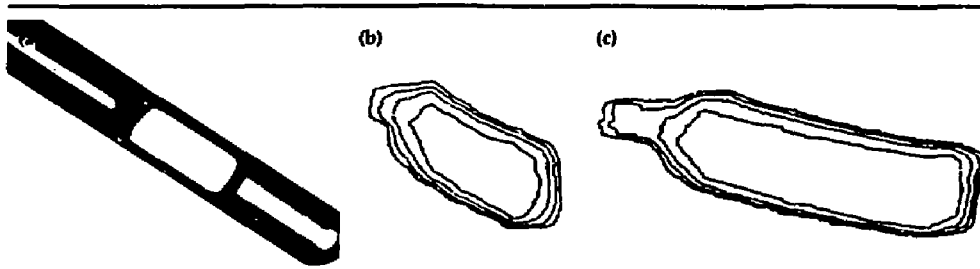


Figure 6. Comparison of image quality obtained at half and full gradient field strengths. (a) 3-mm-diam glass flask used in the experiment. (b) Image obtained at half-maximum strength. (c) Image obtained at full strength. Note that an accurate image of water within the flask is obtained only at full strength.

cant mobile water; hence, no image was obtained. A spectral analysis of one sample revealed no mobile water line but did reveal a very broad line (10 kHz) usually associated with bound hydrogen nuclei. This finding leads us to believe that the water molecules in the minute bubbles are weakly bound to the material walls and are not free to tumble as a liquid, thereby giving solid-like spectroscopic lines.

Unfortunately, at present, the NMR imaging of bound nuclei is not possible with our system. This restriction is primarily a result of limitations in the strength of the gradient fields, thus limiting the amount of change possible in the resonant frequency per unit distance in the gradient direction and making impossible the spatial resolution of nuclei with large line widths.

In a second group of samples supplied by the Nondestructive Evaluation Group, we imaged water in two holes (one of 1-mm-diam and one of 2-mm-diam) in graphite epoxy composite cylinders. An image of such a sample is shown in Fig. 8. The mobile water is easily imaged in these

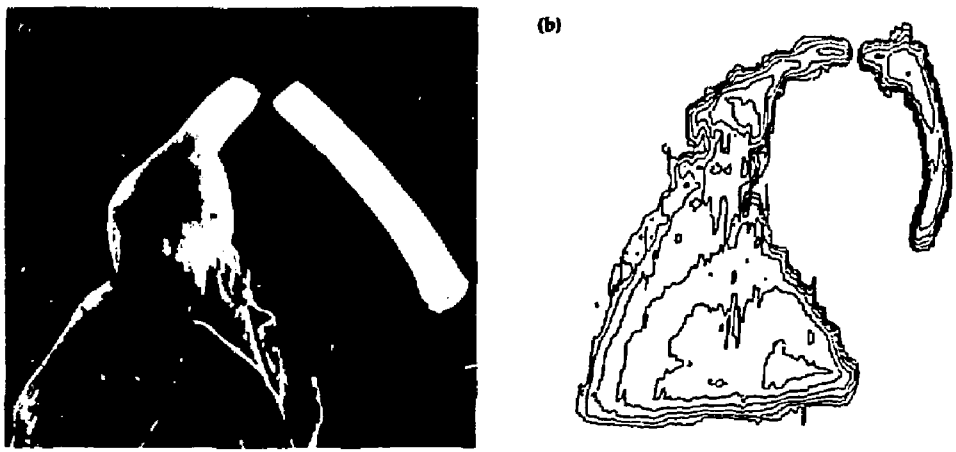


Figure 7. Imaging experiment of a peapod, representing a more complex water distribution. (a) Peapod with broken stem. (b) Image obtained. Note that the break in the stem is preserved.

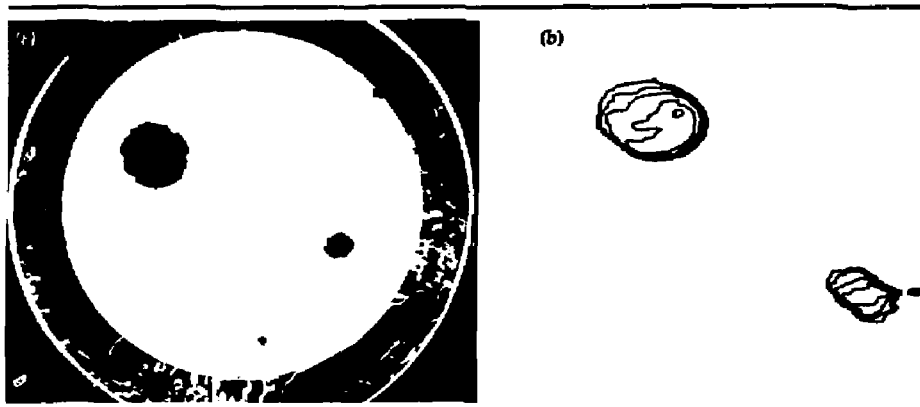


Figure 8. Image of water contained in two holes of epoxy graphite. Note that the relative difference in hole sizes is preserved.

samples, and the relative difference in hole diameters is preserved. The distortion of the circular shape is caused by slight errors in the calibration of the two gradient fields. This image shows that our resolution is significantly better than 1 mm.

A series of experiments was also conducted on welded tuff rock from Yucca Mountain, a site of interest for high-level nuclear waste storage. Several small core samples were obtained and boiled or immersed in water at 60 psi for one week to ensure water absorption so that we could determine the water distribution inside each sample. The rock samples were found to be highly impermeable, with only one sample absorbing a significant amount of water (10%). When the hydrogen NMR spectrum of this sample was obtained, it indicated that a relatively mobile type of water was present. (Two types of bound water were detected in the spectra of the impermeable samples.) The 13-kHz line widths of the two types of bound water were too broad for our gradient fields to resolve. The relatively mobile type of water had a line width of 3 kHz, indicating a loose bonding. Attempts to image this sample were not successful because the line width was still too broad for the gradient fields to resolve.

Future Work

Other improvements under way in the NMR imaging hardware and software include the development of a 512-by-512 pixel, 256-color image display system and the implementation of a projection reconstruction algorithm for image processing. The LLNL Chemistry Department will continue funding this development in FY 87.

Nuclear magnetic resonance imaging applications are important for research in Laboratory programs. Future NMR imaging research at LLNL should proceed along two lines: first, researchers in projects and programs with specific imaging problems toward which the current Sandia instrument (or an outside instrument for which we can contract time) can be applied need to fund continuing efforts. Second, a long-term research effort, in which the feasibility of using line narrowing techniques would be investigated, should be instituted. This effort could yield long-term breakthroughs in the nondestructive evaluation of materials and fundamental insights into organic and inorganic chemical processes. Instituting such a long-term research effort requires the acquisition of a general-purpose high-resolution solids research spectrometer with imaging capability. Unfortunately, the cost of such an instrument is high ($500 000) but, if also used as a spectrometer to satisfy present programmatic requirements, its purchase could be justified as a multidisciplinary capital-equipment expenditure. The Chemistry Department has approved the purpose of a spectrometer that could be adapted commercially for imaging. The imaging accessory costs $100 000 and has not yet been approved.

Outside Contacts

Our major contact with the outside engineering community was with the Health Science Center of the University of Texas at Dallas. Evelyn Babcock and co-workers had modified a spectrometer very similar to ours for imaging. Their information and support was invaluable.

We invited two NMR experts to consult and lecture at LLNL. Brian Suits of the Physics Department of Michigan Technical University consulted on NMR imaging reconstruction algorithms and Mike Stoll, who originally designed the Sandia spectrometer, consulted on the details of that instrument.

Contacts were also made with the NMR imaging researchers at Nalorac Corp. in Concord, CA, the General Electric Group in Fremont, CA, the Bruker Corp. in Massachusetts, and Auburn International in Pleasanton, CA.

Financial Status

The expenditure of Engineering Research funds on this project for FY 86 was about $51 000. Although a larger sum was allotted,

joint funding through the Chemistry Department and the Earth Sciences Department allowed us to accomplish our FY 86 work with less Engineering Research funding.

1. P. Lauterbur, "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance," *Nature* **242**, 190–191 (1973).
2. R. Damadian, "Tumor Detection by Nuclear Magnetic Resonance," *Science* **171** 1151 (1971).
3. W. P. Rothwell and P. P. Gentempo, "Nonmedical Applications of NMR Imaging," Bruker Corporation Annual Report, MA (1985).
4. G. C. Chingas, J. B. Miller, and A. N. Garroway, "NMR Images of Solids," *J. Magn. Res.* **66**, 530–535 (1986).
5. R. J. Gummerson, C. Hall, and W. D. Hoff, "Unsaturated Water Flow within Porous Materials Observed by NMR Imaging," *Nature* **281**, 56–57 (1979).
6. S. Emid and J. H. N. Creyghton, "High Resolution NMR Imaging in Solids," *Physica* **128B**, 81–83 (1985).

Quantitative Radiography

J. M. Brase, H. E. Martz, and K. E. Waltjen
Nondestructive Evaluation Section
Electronics Engineering Department

R. L. Hurd and M. G. Wieting
Nuclear Energy Systems Division
Electronics Engineering Department

Radiographic techniques have been used in nondestructive evaluation primarily to develop qualitative information (i.e., defect detection). In this project, we are attempting to apply and extend the techniques developed in medical x-ray imaging, particularly computed tomography (CT), to develop quantitative information (both spatial dimensions and material quantities) on the three-dimensional (3D) structure of solids. Our accomplishments in FY 86 include (1) improvements in experimental equipment—an improved microfocus system that we believe will give 20-μm resolution and has potential for increased imaging speed, and (2) development of a simple new technique for displaying 3D images so as to clearly show the structure of the object. We have also begun image reconstruction and data analysis for a series of synchrotron CT experiments conducted by LLNL's Chemistry Department. Our future work will extend our current investigations of (1) the use of *a priori* knowledge to improve our imaging systems, (2) discrimination of chemical elementals via 3D imaging, and (3) development of high speed 3D systems.

Introduction

Traditionally nondestructive evaluation (NDE) techniques such as radiography and ultrasonics have been best at furnishing qualitative information (i.e., defect detection) rather than exact measurement of defect parameters such as dimensions, orientation, and material properties. Our goal in this project is to develop an understanding of the ways radiation measurement techniques can be used to determine the three-dimensional (3D) structure of solids. We need to measure both spatial quantities (e.g., position, size, and orientation of cracks) and material quantities (e.g., density and chemical composition).

Our approach has been to apply and extend the techniques developed in medical x-ray imaging, particularly computed tomography (CT), to the different problems encountered in industrial imaging.

Most x-ray procedures have employed the simple projection technique shown in Fig. 1(a). The x rays are emitted from the source (usually an x-ray tube), travel through and interact with the object of interest (represented by the cylinder in the figure), and finally reach the detector (usually photographic film), where their locations and intensities are recorded. This technique is basically the same as that used by Roentgen, who discovered x rays in 1895. The disadvantage of this technique is that the 3D information on the pattern of radiation attenuation in the object of interest is compressed or integrated along the ray paths into a 2D image. Thus, if two important features of the object overlap along these paths, they will also overlap on the image, making it impossible to discriminate important diagnostic features or to identify the material characteristics at a specific point in space.

Computed tomography (CT) was developed as a means of retrieving the 3D information that is lost in standard radiographic imaging. In CT, we acquire radiographic projection images from a set of different angles, then mathematically combine those images in a computer to recreate

the 3D attenuation distribution. This technique has revolutionized medical imaging during the last decade and promises to do the same for industrial imaging.

Most development of CT has been in the medical field. Spatial resolutions have increased from 3 to 0.5 mm and imaging times have decreased from 5 min to a few seconds. As the cost of computers decreases and the need for quality and reliability of industrial products increases, CT is now becoming an important NDE tool. However, the industrial environment imposes a new set of technical demands, particularly for high resolution. To be generally useful in material evaluation, spatial resolutions on the order of a few microns are necessary.

Technical Status

Current Practices and Problems

Reconstructing the data from an entire object can take hours of computer time. Because in NDE evaluations we are usually interested in relatively small areas of the object, we can select single cross-sections to

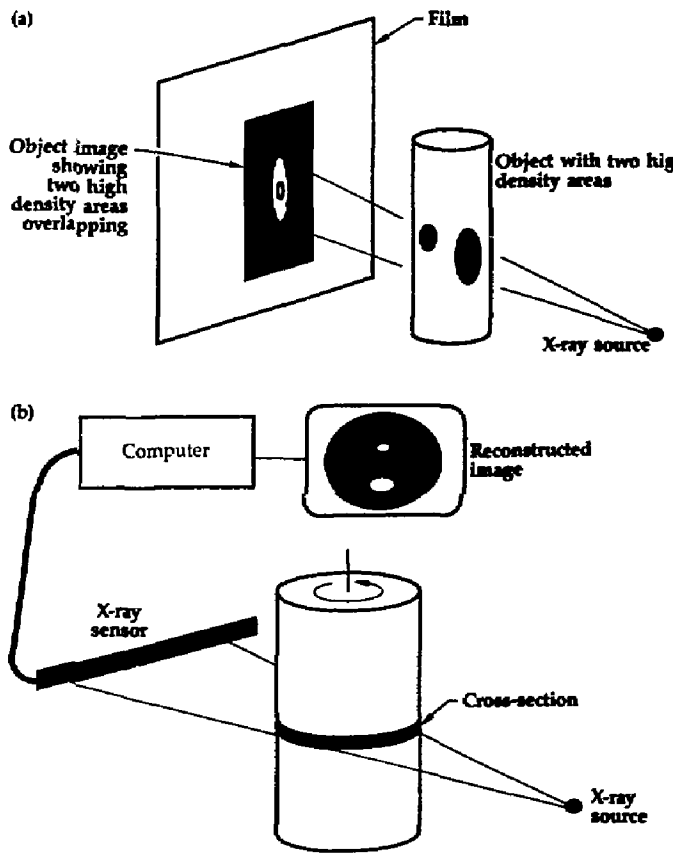


Figure 1. Comparison between standard radiography and CT. (a) In standard radiography, the image is formed by projecting rays from the source to the image plane through the object of interest. Therefore, if more than one object lies along a ray, they will overlap in the final image. This can cause serious problems for interpretation of the image. (b) In CT, an image of a cross-section of the object is produced after computer reconstruction of multiple projections.

form our reconstructions, as shown in Fig. 1(b). Later we can stack these individual slices together to form a 3D image, if desired. In the following, we will generally be talking about scanning and reconstructing single slices, as performed in our project.

Imaging Geometry. Figure 2 illustrates one geometry we use in CT, showing the relation between the chosen angle (θ_i) and the line of projection (D). A projection is defined as the one-dimensional set of data formed by sampling the x-ray intensity at a series of points (P_1, P_2, . . . , P_N) along a line (D). Projections are acquired at a set of discrete angles (θ_1, . . . , θ_M). Typically we record from 300 to 500 points per projection (N) and 90 to 180 projections (M) over 180°.

The particular geometry shown is called "fan-beam." Another common arrangement is "parallel beam," in which the source is collimated down to a pencil beam and a single collimated detector is used. For each angle θ, both the source and detector are translated along line D and data are recorded at the points shown. The parallel beam geometry has the advantage of simple reconstruction techniques but the disadvantage of generally longer data acquisition times, since it cannot use detector arrays. Several other geometries have been used in commercial systems to further speed the scans and simplify the system mechanics.[1]

Radiation Sources for CT. We use three types of x-ray sources for CT: standard x-ray tubes, radioisotopes, and synchrotron radiation. X-ray tubes and radioisotopes are the most common sources but both have significant drawbacks for CT. Tube sources emit a polychromatic beam, which makes reconstruction more difficult, while isotopes have very low intensity.

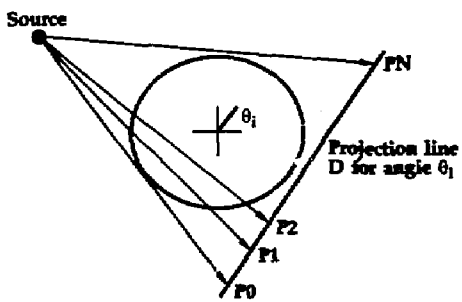


Figure 2. The imaging geometry for CT. The x-ray intensity is measured at the points P_i. The entire set is taken for each angular position θ_i. The setup shown is for the fan-beam geometry we most commonly use.

Synchrotron radiation is produced by particle storage rings such as those at the Stanford Linear Accelerator (SLAC). When the paths of the high energy charged particles (electrons in the case of SLAC) are bent by magnetic fields to keep them in the ring, they emit wide-band radiation. The spectrum ranges from visible light up to medium energy x rays of approximately 60 keV. The major characteristics of this source are its high intensity and its naturally good collimation. The intensity is high enough that we can separate out a very narrow spectral band and still have intensities three orders of magnitude higher than those of most x-ray tubes. The disadvantages of synchrotron sources are their high cost and low availability. We expect that this situation will rapidly change as their utility in x-ray applications becomes apparent.

X-Ray Interactions with Matter. X rays interact differently with materials, depending on the energy of the x-ray photon involved. Therefore, we prefer, whenever possible, to use single energy (monochromatic) radiation (such as that from radioisotopes or synchrotrons), since it greatly simplifies image reconstruction and allows more detailed quantitative analysis of the materials inspected.

The interaction is quantified by the linear attenuation coefficient $\mu(e)$, which is a function both of the element and of the x-ray energy e. If ρ is the probability that a photon of energy e will pass through a unit thickness of the given element, then the linear attenuation coefficient is defined as

$$\mu(e) = - \ln \rho \quad .$$

The attenuation coefficient is measured in terms of inverse length. For example, the $\mu(e)$ for water at e = 71 keV is 0.19 cm^{-1}. In everything to follow, our measurements will be of relative linear attenuations, which are defined as $\mu(e)^t - \mu(e)^c$, where t is the test

material and c is a calibration material, which in our case is normally air. Since μ is a function of the energy, the use of polychromatic sources complicates our measurements of the spatial distribution of attenuation. This will be further discussed in later sections.

Calculation of Ray Sums. When we do a scan, what we actually measure at each point p_i is proportional to the total number of photons arriving at the detector for some constant period of time. What we really want is an estimate of the line integral along the ray R_i of the linear attenuation $\mu(x,y,e)$:

$$m_i = \int_R \mu\,(x,y,e)\,dz \quad ,$$

where m_i is called the monochromatic ray sum for ray R_i and $\mu(x,y,e)$ is the relative linear attenuation coefficient at point (x,y) in the object. It is easy to show that

$$m_i = -\ln(A_m/C_m) \quad ,$$

where A_m is the x-ray intensity at p_i measured through the object, and C_m is a calibration intensity measured at p_i with the object absent. This relationship holds only for monochromatic measurements as denoted by the subscript m on A and C.

As we noted above, however, our measurements are very often made with polychromatic sources. In this case, we calculate the polychromatic ray sum p_i in the same way:

$$p_i = -\ln(A_p/C_p) \quad .$$

In general, p is not the same as m and, in fact, p does not even uniquely determine m. However, it appears we can get useful reconstructions based on p and we are developing means to better approximate m even with polychromatic sources.

Because the x-ray sources often vary significantly in intensity during the course of a scan, we always establish a reference detector that looks directly at the source at all times. The measurements A_p and C_p are normalized to this detector before calculation of the ray sum p_m to correct for source strength variations.

Reconstruction Algorithms. Computed tomography is fundamentally different from most imaging processes in that the data must be heavily processed by a computer before the image can be formed. The reconstruction process uses the set of projection data $p(l,\theta)$, which the computer has stored during the scan to recreate the two-dimensional distribution of radiation attenuation $\mu(x,y)$. The process must take into account the geometric setup of the system, characteristics of the source and detector, and any available *a priori* information about the object being inspected.

If we let the attenuation coefficient distribution be written in terms of polar coordinates $\mu(r,\phi)$, then Radon's theorem gives us an expression for μ in terms of the infinite set of projections $p(l,\theta)$:

$$\mu\,(r,\phi) = \frac{1}{2\pi^2}\int_0^{\pi}\int_E^E \frac{1}{r\cos(\theta-\phi)-1} \times p_l(l,\theta)\,dl\,d\theta \quad ,$$

where $p_l(l,\theta)$ is the partial derivative of $m(l,\theta)$ with respect to l, and E is the size of the reconstruction area. There are several problems with using this formula directly. First, it requires an infinite set of projections to assure accuracy of the reconstruction. Obviously, the best we can do is a finite subset of the projections. Second, the formula is very sensitive to noise such as the statistical fluctuations of the ray sums. Last, the formula does not give us an algorithm for computer implementation. An efficient, discrete approximation must be found with good noise sensitivity.

Computed tomography reconstruction algorithms fall into two basic groups: the transform techniques and the algebraic techniques. The transform techniques are based on Radon's theorem (Ref. 1). The algorithms derived from it must approximate its four parts:

1. A partial derivative with respect to the coordinate along the projection line.
2. A Hilbert transform.
3. Backprojection.
4. Renormalization by $\frac{1}{2}\pi$.

A transform technique called backprojection is the most common reconstruction technique. The partial derivative and Hilbert transform are approximated by a single filtering operation on the projection data. When these filtered projections are backprojected, a very good approximation of the true two-dimensional distribution is formed. This technique is popular because it is easy to implement in either software or hardware and is relatively fast. It produces very high quality images as long as the number of projections is high—close to the number of points per projection.

The algebraic techniques of reconstruction are fundamentally different in that they are completely discrete. Transform techniques treat space as continuous and are converted to discrete approximations only as a last implementation step. In the case of algebraic techniques, though, the model used is discrete from the beginning. If our reconstruction area is $N \times N$ and we represent the attenuation values of the discrete pixels ordered row-wise as x_j, then the unknown image can be written as a column vector $\mathbf{x}$ with N^2 elements. If we define $r_{i,j}$ as the length of the intersection of the jth pixel with the line of the ith projection ray, then we can describe the imaging process as

$$y_i \approx \sum_{j=0}^{N^2-1} r_{i,j} x_j \quad ,$$

where y_i is the ith ray sum (also row ordered), or in vector form,

$$\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{e} \quad ,$$

where $\mathbf{R}$ is the $I \times N^2$ matrix with elements $r_{i,j}$ and is known as the projection matrix. I is the total number of ray sums [I = (number of points per projection)(number of projections)]. The vector $\mathbf{e}$ is the error in the earlier approximation. Thus, given the ray sum data $\mathbf{y}$, our problem is to estimate the image data $\mathbf{x}$.

We can approach this problem by assuming both $\mathbf{x}$ and $\mathbf{e}$ to be random vectors with probability density functions $p_x(\mathbf{x})$ and $p_e(\mathbf{e})$, respectively. The Bayesian estimate is found by maximizing

$$p_e(\mathbf{y} - \mathbf{R}\mathbf{x})p_x(\mathbf{x}) \quad .$$

This will give us the optimal estimate in the sense that the $\mathbf{x}$ chosen maximizes both the fit to the data and the *a priori* probability for possible images. Unfortunately, we are required to know both the distributions of measurement errors and the possible images—things we rarely know very well. Also, it may be extremely difficult to solve for the optimal $\mathbf{x}$, given arbitrary p_x and p_e.

A commonly used optimization approach is to find the least squares solution for $\mathbf{x}$ that minimizes

$$\|\mathbf{e}\|^2 = \|\mathbf{y} - \mathbf{R}\mathbf{x}\|^2 \quad .$$

However, $\mathbf{x}$ is in general not uniquely determined by the least squares method and we must apply a second criterion. Also, the least squares solution is simply the one that best fits the data; it includes no information about desirable images. Secondary criteria that can be used include minimum variance or maximum entropy.

Algebraic techniques are usually much slower than transform techniques and, for large amounts of data, do not produce superior reconstruction results. However, they can do a much better job when the number of projects is small; for example, if we had a restricted view of the part and could acquire data only at certain angles. These algorithms may also be very useful if the data acquisition time is much greater than the reconstruction time. For example, if we scan with a weak radioisotope source, the acquisition time can be hours. If we can do a good reconstruction from only a few projections, the radioisotope technique becomes practical.

Research Tasks and Current Status

During FY 86 we have addressed several different research tasks.

A Microfocus CT System. There are four essential elements in any CT system. First is a radiation source, such as those described

previously. Second is the electro-mechanical system for enabling the measurement of projections from different angles. In our case, this is a simple rotational stage that holds the object of interest. Third is the detector system, which measures the intensity of the x rays that penetrate the object. Last is the computer, which acquires the measured data and then reconstructs the cross-sectional image. Figure 3 shows a complete system for CT data acquisition and display.

The microfocus CT system we are developing at LLNL differs from the standard system in several components. The tubes used as radiation sources in most standard systems have electron beams ranging in size from 0.5 to 1.0 mm. This size is a determining factor in the spatial resolution that can be achieved. We use as the source a microfocus x-ray tube with a beam size of approximately 10 μm. Because of this small spot size, we can use geometric enlargement due to beam spreading to increase our resolution. The magnification is given by

$$M = (d_1 + d_2)/d_1 \quad ,$$

where d_1 is the distance from the source to the object and d_2 is the distance from the object to the detector. We believe we can attain approximately 20-μm resolution with this system.

The most common detectors used for CT are various types of scintillating crystals that convert x rays to visible light. In our systems, we use a thin screen that converts the x-ray spatial distribution to a visible light distribution. The light then is amplified by an image intensifier and recorded by a television camera.

The cameras used to convert the light signal to an electrical signal are based on charge coupled devices (CCDs). These solid-state components measure the photon intensity on an array of photosensitive sites and produce an output voltage that represents the charge accumulated in each site sequentially. CCDs work well for tomography because of their wide dynamic range of approximately 10 000:1, their low noise, and spatial and radiometric linearity. The sizes of arrays used are typically 390 × 480 or 780 × 480.

The analog video signal is digitized at video rate (about 15 MHz) by an IP-512 image processor. Because of the high speed required by the video signal, the x-ray intensities are digitized to only 8 bits or 256 discrete levels. Thus, our high speed

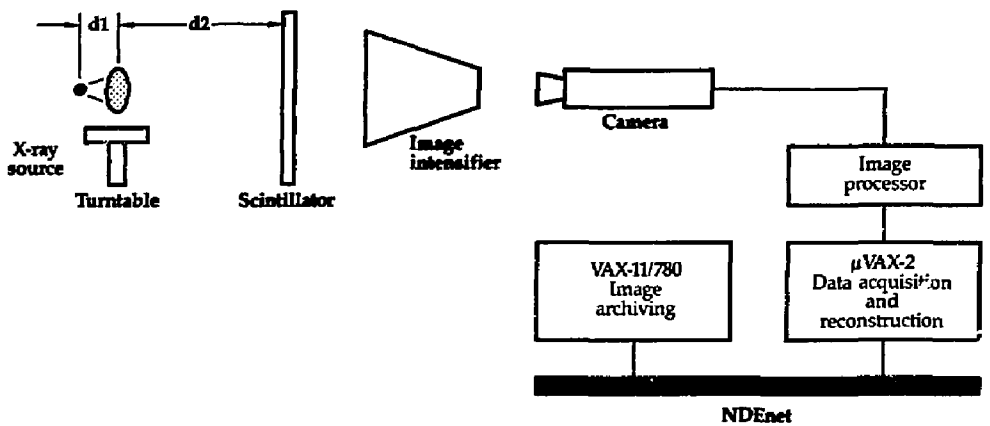


Figure 3. Computed tomography data-acquisition and display system. The microfocus CT system uses a source with an effective spot size of 10 μm. This allows us to use geometric magnification ($M = d_2/d_1$) to increase image spatial resolution. We have demonstrated a resolution of 100 μm.

digitizer loses most of the dynamic range of our CCD detector. One of our first planned upgrades is a controllable integration time camera with a cooled CCD detector and a high resolution, 16-bit, analog-to-digital converter.

The computer in our system is a DEC μVAX-2 running the VMS operating system. It can acquire the data for a single 512 × 512 cross-sectional image (180 projections) in about 5 min and reconstruct the image in about 15 min, using the filtered backprojection technique. This speed is adequate for single-slice imaging; but for complete 3D imaging, we need to speed up the reconstruction. Since it must be done for each slice individually, the total reconstruction time for a 512 × 512 × 512 3D image is 7680 min or 128 hr. Current medical scanners with hardwired reconstruction algorithms reconstruct one slice of this size in about 30 s, and the entire image in about 4 hr. Obviously, these times are still impractical for routine industrial inspections. New ideas in parallel computer architectures are needed to speed this process.

We are currently characterizing the resolution and sensitivity of the microfocus CT system. The resolution is specified in terms of spatial frequency by the modulation transfer function (MTF). This function gives the level of response to a given spatial frequency and is given by the Fourier transform of the system point spread function. Thus, to measure the MTF, we image a point object and take the Fourier transform. Alternatively, we may look at an edge response and calculate the MTF, since we know the spectrum of an edge. Other parameters we are measuring are signal-to-noise ratio and contrast sensitivity.

Three-Dimensional Imaging and Display. Our microfocus CT system is ideal for three-dimensional imaging. The 2D geometry of the fluorescent screen and video camera detector system make the data acquisition time constant whether we are recording the entire object or a single slice. To record multiple slices, we store a rectangular section of the projection image rather than a single line. Since the integration time of the camera (currently 1/30 s) is fixed, the recording time increases only by the small amount necessary to store the additional lines on disk. Each recorded slice is then reconstructed normally.

We have developed a simple new technique for displaying 3D images in such a way that the structure of the object is clearly shown. The multiple slices are displayed in diagonal overlapping areas of a graphics window, as shown in Fig. 4. The normal image color map is modified so that the original image values (0 to 255) map into four colors: red, green, yellow, and cyan. Depth in the object is indicated by the lightness of the color displayed (Fig. 5). A pixel in a back slice with value 100 will be a very dark green, while a pixel with the same value in a front slice will be a very light green. The combination of the diagonal overlap and the lightness variation creates a shaded 3D image that can be drawn very quickly. We also must define a transparency level. A pixel with a value less than this limit will not be drawn. This allows us to see through very low valued material such as air.

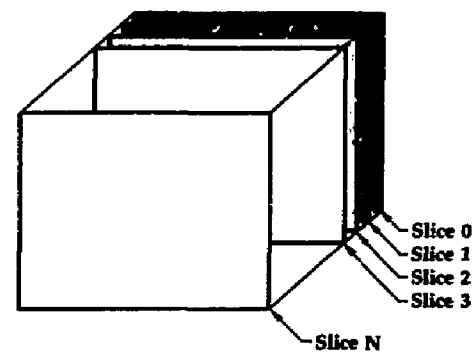


Figure 4. Three-dimensional displays produced by displaying each slice individually while shifting the origin of each display diagonally to produce depth. The color shades are progressively made lighter from back to front. At each pixel of each slice a decision must be made, based on the transparency level, whether or not to display the pixel.

This display also allows us to easily remove sections of the object by simply setting coordinate boundaries that determine the limits of the display. Using this, we can look at internal details of the object that otherwise would be obscured by the object surface. We can also do simple right-angle rotations by changing the order in which we address the 3D image data.

The major drawback to this technique is that it does not create perspective in the image. If the object is very deep, it will appear distorted, and it may be necessary to go to a more complex technique to display the object image. The technique works best if the number of slices is much less than the size of the reconstruction array; for example, for a 256 × 256 reconstruction array, 50 to 60 slices will work very well, but if we use over 100 slices, the image starts to show obvious distortion. The display also loses CT number resolution, compared to our single slice displays, which can show 256 colors. Since we need to show approximately 64 lightnesses of each color for depth information, we are left with only four to show ranges of CT number. The best strategy appears to be to use the 3D display for overall form and the single slice images for detailed attenuation information.

Figure 5. Use of multiple slices to form a 3D image of an object. Shown here are 40 slices of an aluminum piston. Each slice was acquired at 150 keV. A quarter of the object image was removed to show interior details. In our technique, depth is represented by the darkness of each color, while different hues represent different densities.

Synchrotron CT. Our third basic task has been participation in a series of synchrotron CT experiments carried out by J. Kinney of the LLNL Chemistry Department. In these tests, carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), we have been responsible for image reconstruction and data analysis. This effort has just begun and we expect it to carry on in FY 87 and beyond.[2]

The synchrotron CT system looks much like our microfocus CT system in that it consists of the same basic components. The differences lie mainly in the nature of the x-ray source and the detector system used. Since the characteristics of synchrotron radiation have been discussed above, we will discuss here the detector system and the new techniques it has made possible.

The detector used in these experiments consists of a thin, glass plate coated with a P46 rare earth phosphor for converting x rays to visible light and a TI-4849 virtual-phase CCD detector that is optically coupled to the phosphor. The CCD has a 390 × 580 pixel format, each pixel being 33 μm square. It is cooled to −60°C to reduce the rms noise level to around 10 electrons, giving a dynamic range of about 10 000:1. The CCD can look directly at the scintillator without an image intensifier (as in the microfocus system) because of the high intensity of the source. The output of the CCD is digitized to 14 bits and sent to a μVAX-2 for storage and later reconstruction.

An example of a reconstructed image from this system is shown in Fig. 6. The object imaged consists of two thistle tubes, one filled with copper oxide and the other with nickel oxide. The images shown are reconstructions for 36 projections. Image (a) was recorded at 8800 eV, which is below the copper absorption edge, while image (b) was recorded at 9100 eV, which is above the edge. We see clearly that the image of the copper-filled tube has changed dramatically, while that of the nickel-filled tube has changed very little. Thus, by subtracting

these two images, we should produce an element-specific image containing only the copper. We are currently working on software to allow us to correct positioning errors by registering the two images so that they can be properly subtracted.

Future Work

Our planned work continues in the same three areas as our current work: the use of *a priori* knowledge to improve our imaging systems, elemental discrimination, and high speed 3D systems.

As mentioned before, we often have a substantial amount of knowledge about our object, such as physical geometry and material characteristics. We are looking for new ways to use this detailed knowledge to improve the quality of our images and increase speed by decreasing the amount of required data. For example, by combining the data taken from two different source-to-object distances, we can estimate the scatter contributions and remove it. This reduces one of the primary causes of error for our uncollimated CT systems. Also if we can characterize the effects of the x-ray imaging system, we can use image processing techniques to improve spatial resolution.

For elemental discrimination we are taking two approaches. First, we will continue our series of experiments in synchrotron CT at Stanford. Using the tunability of the synchrotron source to see the fine variations in x-ray attenuation, we expect to form 3D images not only of element distributions but of chemical state distribution. It would then be possible, in inspecting an assembly, to use CT to isolate any areas of corrosion. The second approach is to do elemental imaging using a standard x-ray tube source. By recording several images at different tube voltages and statistically analyzing the resulting image data distributions, we may be able to separate elements. This would make it unnecessary to go to a synchrotron source for many applications.

We noted above that the routine use of 3D CT will require the use of new parallel architectures to achieve reasonable speed in

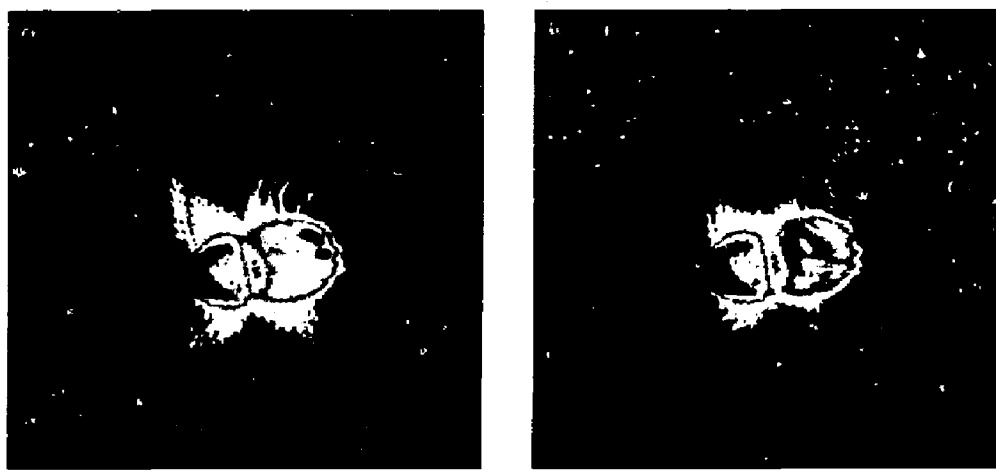


Figure 6. Reconstructed, element-specific image using synchrotron CT. In (a) we show a 36-projection scan taken at 8800 eV at the Stanford Synchrotron Radiation Laboratory. The object consists of two thistle tubes, one containing copper oxide and the other nickel oxide. In (b) the reconstruction of the projections at 9100 eV is shown. The copper image has changed due to moving across copper's absorption edge. The nickel oxide image has remained constant, thus enabling us to distinguish the two elements.

image reconstruction. We are currently developing a systolic algorithm that will run on a 64-processor array.

Beyond the problem of producing meaningful 3D images quickly is the problem of interpreting them. Evaluation of thousands of images looking for small features is a tiring, error-prone job. Automating this process will require the latest techniques in computer vision and artificial intelligence. This job is just beginning.

Outside Contacts

As noted above, we have collaborated with workers from the LLNL Chemistry Department in a series of synchrotron CT experiments. During the next fiscal year we anticipate much broader contacts with the engineering community since our experiment will be more fully developed.

Financial Status

The total expenditure for this project in FY 86 was $69 000.

1. G. T. Herman, *Image Reconstructions from Projections* (Academic Press, New York, 1980).
2. J. H. Kinney, Q. C. Johnson, U. Bonse, R. Nusshardt, and M. C. Nichols, "The Performance of CCD Array Detectors for Application in High-Resolution Tomography," Society of Photo-Optical Instrumentation Engineers, *X-ray Imaging II*, paper 691 (in progress).

Research in Quantitative Ultrasonic Nondestructive Evaluation

G. A. Clark
Engineering Research Division
Electronics Engineering Department

D. K. Lewis
Nondestructive Evaluation Section
Engineering Sciences Division

In quantitative nondestructive evaluation, we conduct laboratory experiments using ultrasonic signals to interrogate and measure the physical properties of materials and interpret the data in meaningful terms. In FY 86 we continued to improve our ability to interpret signals acquired with the ultrasonic test bed and the acoustic microscope.

We made progress in a number of areas of this research. Specifically, we significantly improved the range resolution of A- and B-scans by improving system-identification and spectrum-extrapolation algorithms; used wave propagation models and parameter-estimation techniques to estimate the layer thicknesses and reflection coefficients of layered materials; created an impulsive sound field using a programmable pulser; continued work on modeling the metrology mode of the acoustic microscope with B. D. Cook (University of Houston); and in acoustic material signature (AMS) processing, worked to extend the spectral-domain velocity estimation scheme studied in FY 85.

Our work in AMS modeling and AMS signal processing is scheduled for completion at the end of FY 86. We propose to restructure and rename the resolution and pulse preprocessing tasks in FY 87. At that time, we will focus on enhancing resolution using model-based estimation, the programmable pulser, and a new computer-controlled amplifier. We will begin using special-purpose signal processing hardware to implement our successful algorithms.

Introduction

As discussed in the FY 85 *Engineering Research Annual Report*,[1] the noninterpretability of data acquired using the ultrasonic test bed and the acoustic microscope (both recently developed at LLNL) clearly illustrates the need for advances in quantitative nondestructive evaluation (QNDE). Both instruments consist of mechanical manipulation systems that position transducers and associated electronics that generate and measure ultrasonic pulses. Conventional QNDE displays are based on simplistic signal processing methods of generating, detecting, processing, and displaying information. Therefore, while the digital data acquired from these instruments is rich in information, interpretation is very difficult. In FY 86 we continued to improve our ability to interpret these data.

Using ultrasonic signals to interrogate solid objects can yield a variety of information about inhomogeneous regions within the object's boundary—for example, cracks, bond regions, and voids. It can also provide information about properties of the bulk material, including residual stress, surface texture, and plate thickness. The character of an object is encoded in the ultrasonic field from an external transducer. We can measure the scattered field to infer the nature of the object. Doing so requires the hardware to direct and receive the sound and the tools to deconvolve the properties of the transducer from the received signals. In conventional NDE analog systems, the signals are processed for presentation without removing the transducer characteristics. By applying

digital signal processing techniques to the acquired data, we hope to ascertain the scattered field.

In particular, we defined four tasks:

- Task I: Acoustic Material Signature (AMS) Modeling. Develop a model for the metrology mode of the acoustic microscope.
- Task II: AMS Signal Processing. Estimate surface wave speeds from noisy, imperfect AMS data.
- Task III: A- and B-Scan Resolution Enhancement. Improve system-identification, spectrum-extrapolation, and model-based estimation algorithms.
- Task IV: Pulse Preshaping. Generate an electrical driving signal for an ultrasonic transducer such that the transducer sound field approximates a narrow pulse.

The team contributing to this effort includes: G. A. Clark (EE Dept./Engineering Research Div.), D. K. Lewis (ME Dept./Engineering Sciences Div./NDE Section), F. L. Barnes (EE Dept./Engineering Research Div.), S. E. Benson (ME Dept./Engineering Sciences Div./NDE Section), K. E. Waltjen (EE Dept./Nuclear Energy Systems Div.), B. D. Cook (University of Houston), L. Adler (Ohio State University), and R. Weglein (independent consultant).

Technical Status

Task I: AMS Modeling

We have pursued our AMS modeling effort through a contract with Professor B. D. Cook at the University of Houston. Cook has proposed an acoustic-optic variation of the acoustic microscope model to prove the validity of a reflection term that is used in most theoretical models [R1,2].* He found that this commonly used reflection model is simplistic and does not agree with reality.

The mathematical model of the acoustic microscope can be reduced to evaluating one equation:[2]

$$v(Z) = \int_0^{\alpha} r|U_1^+(r)|^2 P_1(r)P_2(r)R \times (r/f)\exp\{-j(k_0 Z/f^2)r^2\}dr \quad . \quad (1)$$

The $V(Z)$ referred to at LLNL is $V(Z) = 20\log[|v(Z)|/v_{ref}]$. The terms contributing to the integral are the following:

1. The variable of integration r is the polar radial variable; the focal length of the microscope lens is f, and the wave constant $k_0 = \omega/c$ involves the sound velocity of the liquid couplant.
2. $P_1(r)$ and $P_2(r)$ are pupils of the lens. In the first approximation, they are unity. Aberrations, finite lens size, and other limitations can be introduced through them. For example, large numerical aperture effects studied by Sheppard and Wilson[3] have been introduced using them. In general, the functions are not known in an exact manner in any experimental configuration.
3. $|U_1^+(r)|$ is the sound amplitude distribution at the focal plane of the lens (inside the buffer rod). Based on an assumed distribution at the transducer, Cook has developed a method to calculate this term in the formulation of Gauss-Laguerre polynomials.
4. $R(r/f)$ is the reflection coefficient for plane waves striking the planar liquid-solid interface. The periodicity of the $V(Z)$ curves occurs because of the behavior of this reflection curve.

To test whether the commonly used reflection coefficient was valid, Cook built an acoustic-optic detector [R2]. The results of his investigation indicate that the simplistic model for the reflection coefficient is wrong. The model assumes that the solid occupies a semi-infinite half space, whereas it is usually a thick plate. Consequently, sound that enters the solid is reflected from the far side of the plate and back into the microscope.

Cook is now gathering reflection routines for plates for inclusion with the routine that evaluates the integral in Eq. (1). Once he modifies this routine, he will compare the results with those obtained using the acoustic microscope. This comparison is scheduled for FY 87.

* Reference citations prefixed with the letter R denote documents partially or fully supported by this project. These documents are listed at the end of this article.

Task II: AMS Processing

In FY 85 we showed that under some conditions, the Rayleigh wave velocity can be estimated from the spatial frequency spectrum of the AMS [R3]. For simple cases, in which the AMS contains only one spatial frequency, the spectral method did not work significantly better than the space-domain method. We hypothesized that the benefit of the spectral method will be realized when we have unknown material samples containing several Rayleigh velocities (several spectral peaks), because in this case, the spatial method fails.

In FY 86, to test our hypothesis, we worked with R. Weglein, an independent consultant, to construct experiments for which the correct solution was known *a priori*. We discovered that designing such an experiment is extremely difficult but took the first steps to do so. Weglein discusses the results of some preliminary experiments in a separate report [R4]. Our results are inconclusive, and the problem will require more study. We have not yet decided to continue research in this area but will make a decision in early FY 87.

As part of our work with the acoustic microscope and the AMS, we have been studying the problem of detecting and characterizing disbonds in a graphite epoxy composite material. The objective of this work is to characterize laminates and detect delaminations. The sample was to have been a sheet of 20 parallel uniaxial layers with an embedded delamination dimension of 1 in. In fact, the sample turned out to be biaxial and, worse, composed of a weave of carbon fiber bundles, as shown in Fig. 1. We modeled the sample as a two-dimensional single crystal. The symbols in brackets denote crystal direction.

Using the acoustic microscope, we were able to image the delamination from either side of the sample—that is, through 18 layers or through 2 layers of the fiber-resin composite. The signal from normal incidence showed good amplitude of the top and bottom of the slab and a ring-down tail between them. This mid signal does not seem to be from the individual layers, but rather an artifact of the texture of the nonsmooth surface. Work is proceeding to characterize this surface with a view toward removing its effects.

Because any reflected signal from small delaminations or porosity in the composite matrix can be obscured by this signal, correcting for surface effects should increase the volumetric resolution considerably. A true uniaxial sample (like that shown in Fig. 2) is now being sought, and methods of smoothing the surface of the present sample are being considered, among them thin and thick cover layers of epoxy. Should a uniaxial sample be found, the AMS can then be used to determine the surface wave speed as a function of angle relative to the [1,0,0] direction.

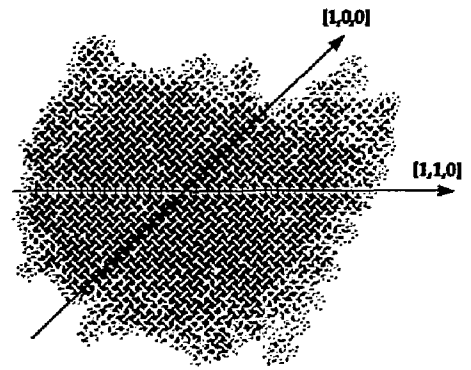


Figure 1. A woven layer of composite fiber, with crystal model axes shown.

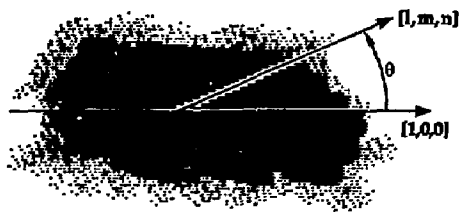


Figure 2. Ideal uniaxial composite.

Task III: A- and B-scan Resolution Enhancement

Our immediate goal is to estimate the impulse response of acoustic scatterers (e.g., flaws) in materials. By doing so, we hope to detect, locate, and estimate the size of the scatterers. We test materials using ultrasonic transducers in a pulse-echo mode, as shown in Fig. 3. The time-domain waveform produced is called an A-scan. If we move the transducer in one dimension, we can collect a family of A-scans to form a two-dimensional image, or B-scan, of a slice through the test sample. If we scan the transducer in two dimensions and record the peak height of the A-scan at a predetermined delay time, we can produce a color-coded image, or C-scan, of the echo magnitude at a particular depth in the part. Figure 3 shows that even though an approximately ideal electrical pulse is applied to the transducer, the sound field produced by the transducer is distorted by the bandpass nature of the transducer and wave propagation paths. For the two-layer problem shown, we measure three wavelets representing reflections from the front surface, middle interface, and back surface, respectively. We are developing methods for removing the distortion from these wavelets.

System-Identification Studies. Our goal is to remove distortion so we can characterize critical flaws and other material properties. The relation below suggests a way to reach this goal. If we assume that we can model the transducer and the wave propagation paths as linear, time-invariant systems, we can write the received signal $y(t)$ as

$$y(t) = u(t) * T_1(t) * P_1(t) * h(t) * P_2(t) * T_2(t) + n(t) \quad , \quad (2)$$

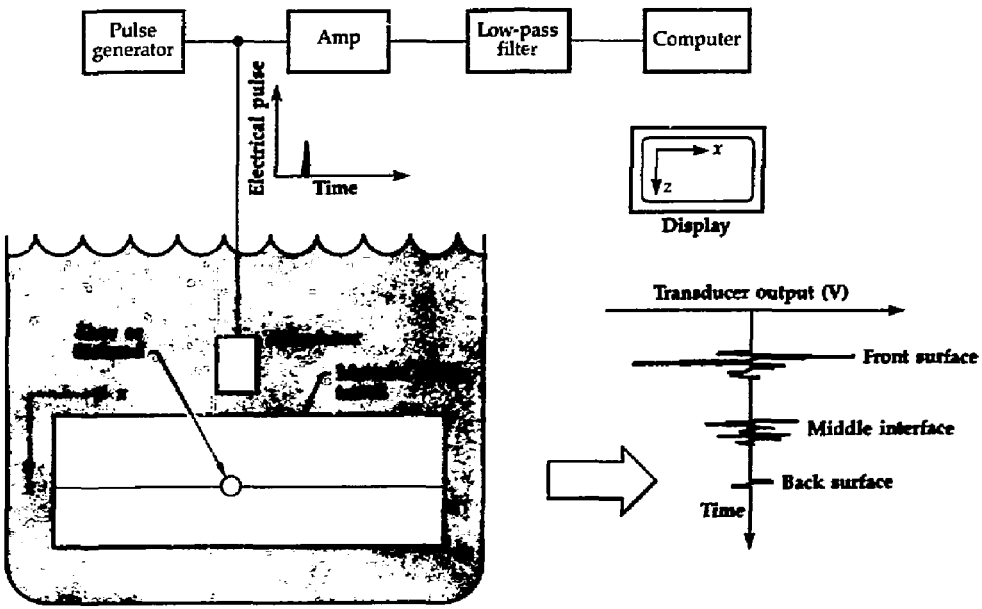


Figure 3. Experimental test setup to measure A- and B-scans. The transducer output for the A-scan shown indicates reflections from the front surface, middle interface, and back surface of the layered material under test.

where $u(t)$ is the electrical impulse driving the transducer, $T_1(t)$ is the forward transducer impulse response, $P_1(t)$ is the forward propagation path impulse response, $h(t)$ is the impulse response of the scatterer of interest (e.g., a flaw), $P_2(t)$ is the return propagation path impulse response, $T_2(t)$ is the backward transducer impulse response, $n(t)$ is additive noise, t denotes time, and $*$ denotes the convolution operation.

Even though we can generate a pulse $u(t)$ that is approximately ideal, the transducer and the propagation paths distort the sound field because they have bandpass frequency responses. The measured signal $y(t)$, then, is far from a true impulse response of the scatterer. Typically, $y(t)$ contains ringing and spreading that severely distort the A- and B-scans, giving them poor resolution; therefore, flaws of interest often go undetected.

If we can find a known scatterer that reflects most of the energy impinging upon it back to the transducer, then we can estimate its impulse response to be the ideal impulse $\delta(t)$ and use it as a reference. Under this condition, we can write the signal measured from the reference as

$$x(t) = u(t) * T_1(t) * P_1(t) * \delta(t) * P_2(t) * T_2(t) + n(t) \quad . \quad (3)$$

The choice of a reference signal can significantly affect the results of processing. Sometimes an appropriate reference may not be available, but in many cases of practical interest, a good reference can be found. We have used corners, front surfaces, and back surfaces for different applications, as discussed in the examples below.

We can now think of the scatterer impulse response $h(t)$ as a linear time-invariant system with input $x(t)$ and output $y(t)$. Thus, our problem is the following: Given the measurements $x(t)$ and $y(t)$, estimate the impulse response $\hat{h}(t)$. This is known as the system-identification problem.[4–11] In principle, the spectrum $\hat{H}(f)$ can be found for the noiseless case by dividing the power spectral densities of the measured signals: that is, $\hat{H}(f) = S_{xy}(f)/S_{xx}(f)$.

The $\hat{h}(t)$ is the inverse transform of $\hat{H}(f)$. In practice, the problem is much more difficult than this, and the solution is much more involved than the simple division of spectral densities. It is an inverse problem which, in general, is ill-posed and ill-conditioned. Our task, then, is to best approximate the solution to this problem by using various forms of regularization methods. By this, we mean that we redefine the problem to remove ill-conditioning and solve the re-defined problem. The main issue is to choose a regularizing scheme that allows us to obtain meaningful solutions.

It is important to understand what we do with $\hat{h}(t)$ once we calculate it. Currently, A- and B-scans consist of the distorted $y(t)$ signals described above. We replace $y(t)$ with $\hat{h}(t)$ when constructing B- and C-scans. This improves our ability to interpret the scans.

We have applied several algorithms to both real and simulated ultrasonic signals to solve this problem. The digital signal processing methods include frequency-domain division, time-domain Wiener identification with regularization, frequency-domain Wiener identification with regularization, and a nonlinear least squares approach. We have also studied several more sophisticated algorithms, as described in [R5,R6].

Bandlimited Spectrum-Extrapolation Studies. Typical ultrasonic transducers used for pulse-echo purposes have a bandpass frequency response that limits the spectrum $X(f)$ of the interrogating signal $x(t)$. Due to noise and the fact that the transducer has such a narrowband response, the spectrum takes on significant values only in some frequency range $(f1,f2)$. This spectral limiting means that the frequency response $\hat{H}(f)$ estimated by a system-identification algorithm is reliable only in the frequency band $(f1,f2)$ (Refs. 12–16). Our goal, then, is to keep $\hat{H}(f)$ only in the region $(f1,f2)$ and to extrapolate $\hat{H}(f)$ outside $(f1,f2)$. The

expected benefit is that the extrapolated impulse response $h_e(t)$ should exhibit improved spatial resolution of pulses.

Several bandlimited signal-extrapolation algorithms have been studied in the literature.[12–16] Most of them iterate between the discrete time domain and the discrete Fourier transform domain with a variation of the method of alternating orthogonal projections. All of the techniques produce good results in the absence of noise but are not very effective when noise is present. We use a variation of the method of alternating orthogonal projections in which a special adaptive algorithm is used to reduce the adverse effects of noise.[12] This regularization technique allows the extrapolation to be completed successfully.

Figure 4 shows the results of applying the regularized system-identification and spectrum-extrapolation algorithms to known layers of Lucite and aluminum. The goal is to improve the temporal resolution of the reflected signals, thereby improving our ability to estimate the layer thickness by measuring the time delay between pulses. We also wish to improve our ability to infer material properties by interpreting the magnitudes and signs of the pulses. The ideal impulse response (lower left in Fig. 4) should be a weighted sum of three delayed impulses, as shown. The pulses represent the reflections from the water-Lucite, Lucite-aluminum, and aluminum-water interfaces, respectively. The raw signal contains broad, ringing wavelets, and the extrapolated impulse response approximates the ideal impulse response, showing a significant improvement in temporal resolution. If the layers were thinner, the pulses would overlap, and the improved resolution could be important in separating the pulses.

Figure 4 shows that the signal processing improves resolution and interpretability for this problem. We have obtained consis-

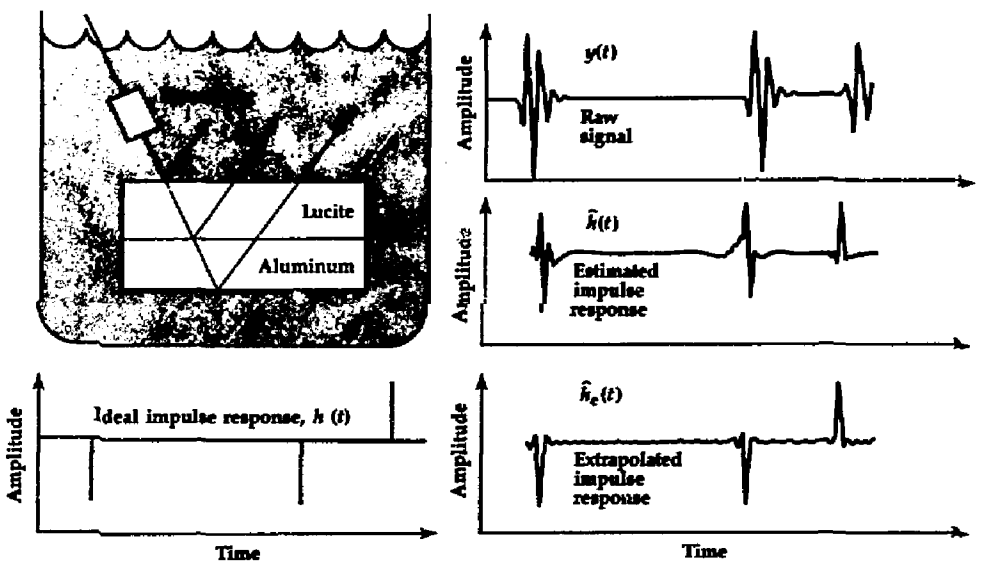


Figure 4. A comparison of the ideal, measured, and processed signals for an experiment to estimate layer properties in Lucite and aluminum. The raw signal is distorted and does not correspond to the ideal signal. The processed signal closely approximates the ideal signal, and is therefore easier to interpret.

tently good results in other experiments. For example, Fig. 5 shows the experimental setup we used in a butt weld underbead experiment. Figures 6 and 7 show the results of this experiment, in which we used an actual programmatic part and actual laboratory signals with our algorithms. The improvement in pulse resolution allows us to more accurately measure weld thickness, which is the prime indicator of weld integrity for this problem.

Model-Based Estimation of Material Parameters. The deconvolution and extrapolation schemes described above have proved successful in solving certain QNDE problems [R3, 5–14]. These schemes comprise a technique that involves estimating the impulse response of a system, then using features of the impulse response to estimate the properties of acoustic scatterers in the material under test. The key point is that the technique does not explicitly use a physical model for the material. Instead, it is a nonparametric modeling approach, in which only the data are used.

The model-based approach described in this subsection is more direct. Here we use *a priori* assumptions or knowledge about the material along with models of wave propagation in the medium. We use these to build an estimator that computes values of the material properties directly, without going through the intermediate step of computing the impulse response. The class of problems we are attacking, however, is more restrictive than the nonparametric system-identification case, in the sense that the model-based approach is applied only to the case of layered media.

Layer problems are important in QNDE at LLNL and elsewhere. Examples include butt welds used in waste repository containers and those in various weapons programs to fuse metals, pinch welds for closing tubes, electronic solid state devices composed of layers of silicon and other materials, diffusion bonds, and the lightweight graphite-epoxy materials used to construct mechanical structures.

While extrapolation and Wiener filter deconvolution are considered to be general signal processing tools, the model-based estimation procedure for layered media has been developed mostly by the reflection seismology community.[17-21] Because the QNDE problem is very similar to that of oil exploration, we can borrow their methods. We are also working on extensions of the general theory, as well as extensions of the applications. The mathematics used is an extension of concepts from the estimation and control theory literature.

The most significant contribution from seismology is the set of functional equations that model the wave propagation in a layered system.[17] We assume a lossless medium being interrogated by a broadband acoustic pulse applied normal to the surface layer. Figure 8 shows the layered medium model we used. The corresponding equation set is

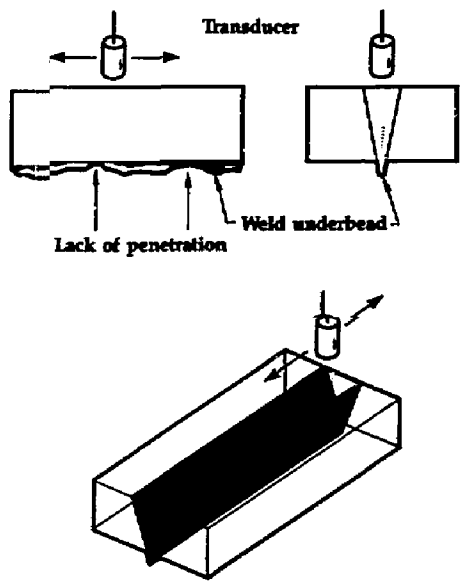


Figure 5. To accurately detect and measure any lack of penetration in the butt weld, we scan along the length of the weld to produce a B-scan image that shows the weld underbead.

$$d_1(k + \tau_1) = -r_0 u_1(k) + (1 + r_0)m(k)$$

$$u_1(k + \tau_1) = r_1 d_1(k) + (1 - r_1)u_2(k)$$

$$d_2(k + \tau_2) = -r_1 u_2(k) + (1 + r_1)d_1(k)$$

$$u(k + \tau_2) = r_2 d_2(k) + (1 - r_2)\, u_3(k)$$

$$d_n(k + \tau_n) = -r\, u_n(k) + (1 + r_{n-1})\, d_n(k)$$

$$u_n(k + \tau_n) = r_n d_n(k)$$

$$y(k) = r_0 m(k) + (1 - r_0)u_1(k) \quad , \qquad (4)$$

where $m(t)$ is the excitation signal (pulse), $d_j(t)$ is the downgoing signal at layer j, u_j is the upgoing signal at layer j, $r_j(t)$ is the reflection coefficient at layer j, and $\tau_j(t)$ is the sound wave travel time for layer j, where $j = 0, 2, \ldots, n$. The materials are assumed lossless, and the excitation signal is assumed to be a plane wavefront normal to the surface.

This equation set, under certain conditions, reduces to discrete state-space form. However, in the practical situation, unlike the discrete state-space form that is discretized in uniform time intervals, the functional equation form is discretized in time intervals that represent the time required for the excitation to travel through the corresponding layer. This implies that

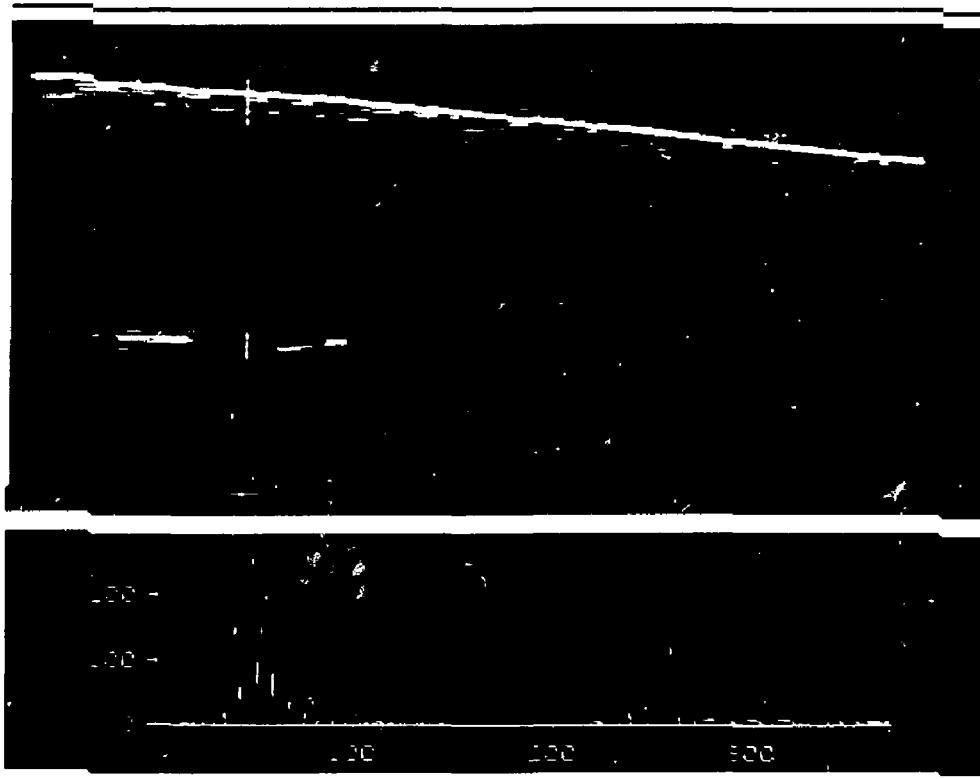


Figure 6. The B-scan of the butt weld from Fig. 5 shows a lack of weld penetration. Note that the image consists of broad lines where we would like to see narrow lines depicting the front and back surfaces of the weld. One rectified A-scan has been selected out of the B-scan and displayed below the B-scan image. It shows the broad, ringing nature of the measured pulses.

each layer travel time is a model parameter, which must be estimated. The objective is to estimate parameters of the layered system recursively: starting from the top layer, estimate τ_1 and r_1; then using τ_1 and r_1 estimate parameters of the next layer, τ_2 and r_2, and continue until the noise level makes the procedure prohibitive. A basic assumption is that additive white Gaussian noise enters the system through the measurement ($z(k)$). The following set of equations relates the set of noisy measurements with layer travel times.

$$\gamma_{j-1} = \begin{cases} 0 & , \quad j = 1 \\ \sum_{i=1}^{j-1} \tau_i & , \quad j > 1 \end{cases}$$

$$Z = |z(t_0), z(t_1), \ldots, z(t_n)|$$

$$Z(\gamma_{j-1}) = |z(t_0 + \gamma_{j-1}), z(t_1 + \gamma_{j-1}), \ldots, z(t_n + \gamma_{j-1})| \quad . \quad (5)$$

Because we do not assume any *a priori* information about the parameters, a reasonable criterion for estimation is maximum likelihood. In the literature,[17-21] it is shown that maximum likelihood estimates of the parameters of layer j may be obtained by first determining the maximum likelihood estimate of the state waveform $u_j(t)$, and then estimating the parameters r_j and τ_j from the state waveform estimate as if the state waveform estimate is the output from layer j.

The connection between the state waveform and the measurement is given by the likelihood equation

$$L_j(r_j,\tau_j) \triangleq \max_{\substack{r_i,\tau_i \\ j<i\leq k}} p\,|Z(\gamma_{j-1})|r_j, \tau_j, \ldots, r_k, \tau_k| \quad ,$$

$$j = 1,2,\ldots \quad (6)$$

The likelihood estimate for the state waveform in layer j is obtained by optimizing the previous likelihood equation with respect to r_j and τ_j and solving for the state waveform. Doing so yields

$$\hat{u}_j(t) = \frac{1}{1 - r_{j-1}} \times |\hat{u}_{j-1}(t + \tau_{j-1}) - r_{j-1}\hat{d}_{j-1}(t)|$$

$$\hat{d}_j(t + \tau_j) = (1 + r_{j-1})\hat{d}_{j-1}(t) - r_{j-1}\hat{u}_j(t) \quad . \quad (7)$$

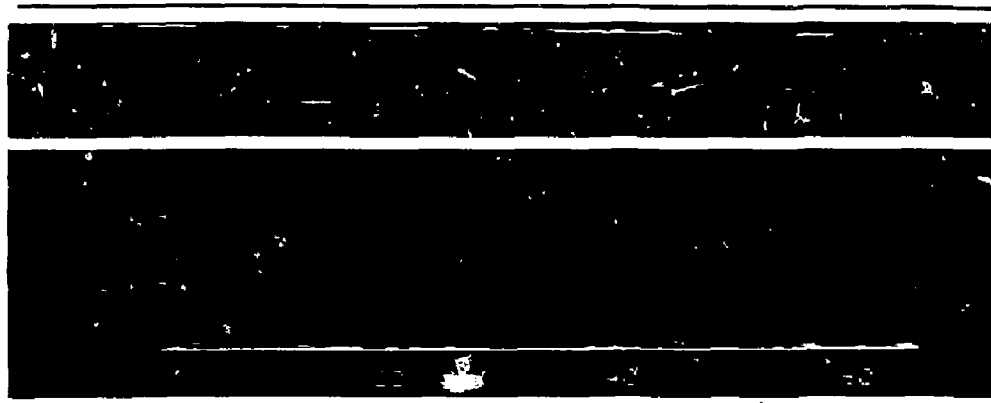


Figure 7. B-scan showing the extrapolated impulse response estimated for the B-scan of Fig. 6. The scan is compressed at the top of the image because of limitations in our display package. Note from the narrow lines in the B-scan and the narrow pulses in the displayed A-scan that the range resolution has been enhanced, making weld depth measurements easier to determine.

The relation of states and parameters is manifest through the likelihood expression

$$L_j(\rho_{j1},\theta_{j1}) = \max_{\rho_{jk},k>1} \left\{ -\int_0^{T_j}\int_0^{T_j} |\hat{u}_j(t) - \sum_{k=1}^{n_j} \rho_{jk} m(t-\theta_{jk})| K_{\hat{u}_j}^{-1}(t-s)|\hat{u}_j(s) - \sum_{k=1}^{n_j} \rho_{jk} m(s-\theta_{jk})|dt\,ds \right\} , \quad (8)$$

where

$$\int_0^{T_j} K_{\hat{u}_j}(t-l)K_{\hat{u}_j}^{-1}(t-s)\,dt = \delta(s-l)$$

$$0 \le s, l \le T \quad . \quad (9)$$

Optimization of this expression leads to likelihood estimates for the parameters, which are

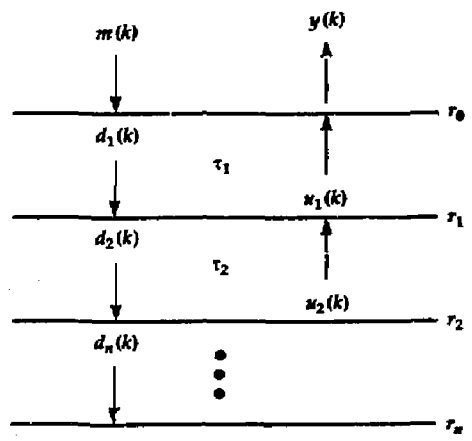


Figure 8. Diagram of a layered medium showing the excitation pulse, downgoing and upgoing waves, reflection coefficients, and layer travel times associated with the idealized model in Eq. (4).

$$\hat{r}_j = \frac{\hat{\rho}_{j1}}{\prod_{k=0}^{j-1}(1+\hat{r}_k)}$$

$$\hat{\tau}_j = \frac{\hat{\theta}_{j1} - \sum_{k=1}^{j-1}\hat{\tau}_k}{2} \quad , j = 1,2,\ldots \quad (10)$$

The model-based parameter estimator uses a two-step process (see Fig. 9). First, we estimate the upward propagating wave u_j and the downward propagating wave d_j for layer j. These are used to estimate the reflection coefficient r_j and the layer travel time τ_j for layer j. This process proceeds recursively through the layers, until no more layers are detected.

We performed a controlled experiment to estimate the reflection coefficient and travel times of layers of fused quartz and water. Figure 10(a) shows the experimental setup, while Fig. 10(b) shows the raw measured signal. The ideal and estimated layer parameters are presented in Table 1. We see that the time delay (layer thickness) estimates are quite satisfactory, but the reflection coefficient estimates are not nearly as good because the model assumes lossless layers, and the actual layers are lossy. In FY 87, we plan to model material losses in an

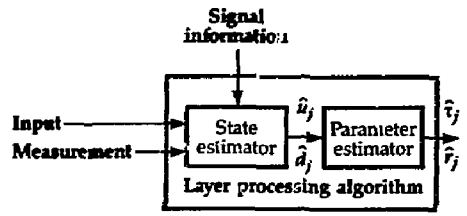


Figure 9. The model-based parameter estimator uses a two-step process. First, the upward propagation wave u_j and the downward propagating wave d_j for layer j are estimated. These are used to estimate the reflection coefficient r_j and the layer travel time τ_j for layer j. This process proceeds recursively through the layers until no more layers are detected.

attempt to improve the reflection coefficient estimates.

Task IV: Pulse Preshaping

After long delays in obtaining a pulser system from one vendor, we ordered a programmable pulse generator from another vendor, which arrived in April 1986. Given a discrete-time waveform (from a computer disk file, for example) this pulser will produce a corresponding analog signal with the same shape and properties. Thus, if we can calculate an appropriate input signal for the pulser, we can use the configuration in Fig. 11 to attempt to produce a sound field measurement that is approximately impulsive when a reference scatterer is being used. Our goal is to minimize the mean square error between the measured waveform and the

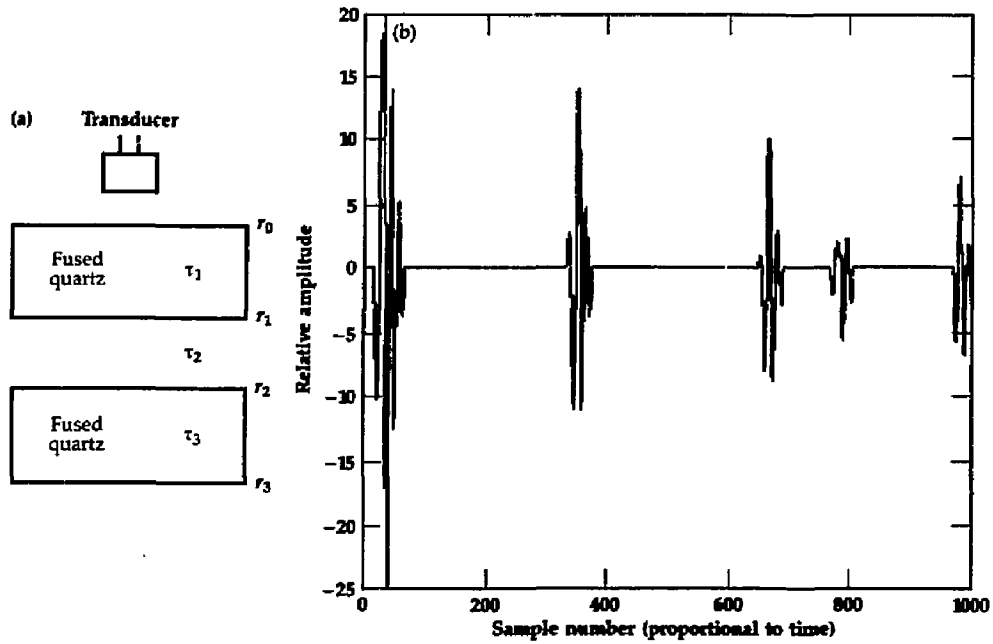


Figure 10. (a) The experimental setup (in a water bath) used to estimate reflection coefficients and layer thicknesses (travel times) for layers of fused quartz and water. (b) The raw signal measured from the layered medium in 10(a).

Table 1. The ideal and estimated reflection coefficients and travel times for the layers of fused quartz and water described in Fig. 10. The symbol j denotes the layer index.

	Reflection coefficient		Travel time	
Layer index (j)	Ideal (r_j)	Estimated ($\hat{r}_j$)	Ideal (τ_j)	Estimated ($\hat{\tau}_j$)
0	0.66	0.66	—	—
1	−0.66	−0.60	158	159
2	0.66	0.31	217.5	217.5
3	−0.66	−0.50	99.5	98

desired waveform (chosen to approximate an ideal pulse). This represents an approach opposite to that of the deconvolution method

We were able to interface the pulser to our minicomputer (a software and hardware effort) and obtain ultrasonic measurements. We performed experiments with known materials, and through off-line computations, we were able to produce a compensated measurement signal that closely resembled an impulse (see Fig. 12), which was our goal. These are preliminary results, however, and in the next year, we hope to generalize the method and expand the applications to more of the actual materials we test in the laboratory.

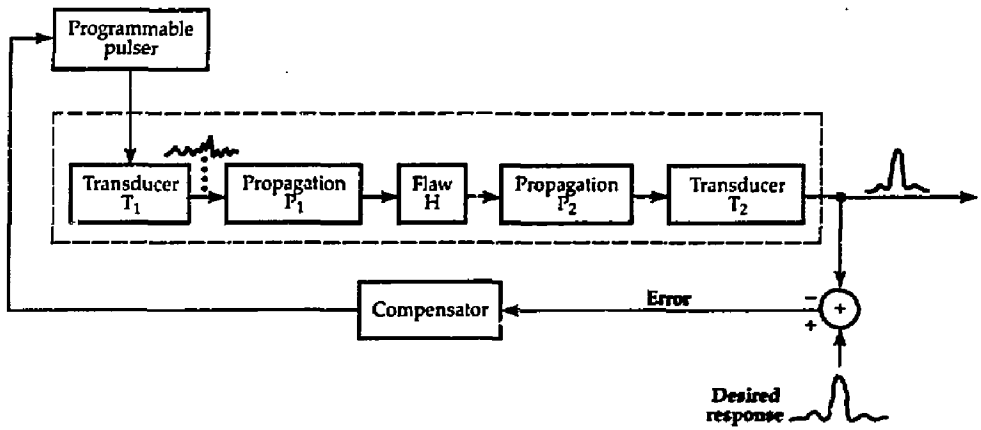


Figure 11. The general pulse preshaping concept uses a programmable pulser to drive the acoustic system under test. Given a discrete-time waveform, the pulser produces the corresponding analog waveform. The goal is to minimize the mean square error between the measured waveform and the desired waveform.

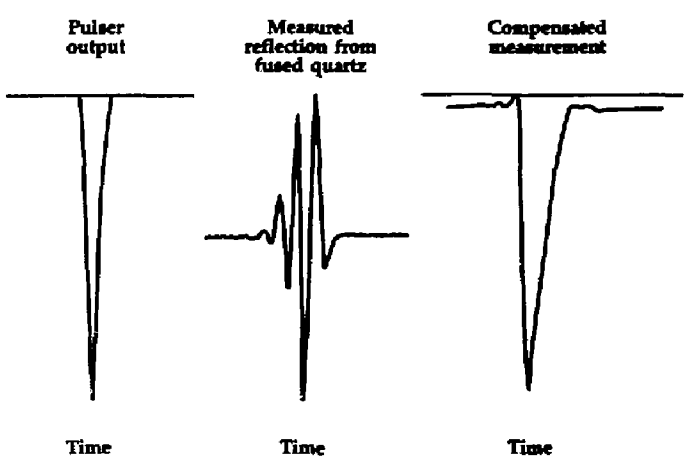


Figure 12. The results of preliminary pulse preshaping experiments are encouraging. We were able to create a sound field that approximates an ideal impulse.

Preshaping the initial pulse has worked very well. We were able to load a Gaussian pulse into a computer file, download that to the waveform generator, and send it to the transducer. To test the accuracy of the resulting mechanical response, Cook suggested and carried out an experiment to determine the system response and the scattering results from a solid elastic sphere.

Cook assumes that the response of a transducer to a unipolar pulse is bipolar due to the mechanics of the problem. Figure 13 shows that if the crystal backing is a perfect absorber, there will still be a rarefaction from the back face in addition to the compression from the front face. Water should behave the same way for nonfinite amplitude waves, so the generated pulse is differentiated once by the transducer, once by the water, and again by the transducer. The calibrating experiment used a single transducer in a pulse-echo arrangement, and the assumption was made that the reflection from a flat, smooth wall of fused silica would not differentiate. Next, a steel ball was used as a target, and once again the pulse was differentiated by the transducer, the water, the water after the scattering, and the transducer. When the signals were integrated, a rounded shape whose time dimensions multiplied by the wavespeed in water were in good agreement with the actual dimensions of the sphere. The preshaping system seems to work, and the next problem is how to determine the optimum initial shape for a given problem.

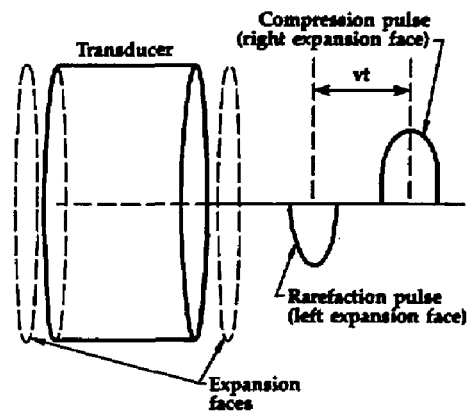


Figure 13. Unipolar input to a perfectly backed piezoelectric crystal results in bipolar output.

Future Work

This is the second year of a proposed three-year research project. Tasks I and II are scheduled for completion at the end of FY 86. We propose to continue Tasks III and IV in FY 87, but we have restructured and renamed them to reflect their changing emphasis, and to avoid confusion with our previous projects.

Task A, Resolution Enhancement. We shall focus on two main areas: (1) In model-based estimation for acoustic scattering, we propose to extend our current algorithms to model the cases of lossy systems, dispersive systems, and non-normal incidence of the driving signal, in order of difficulty. (2) We will use our new programmable pulser in laboratory experiments to produce impulsive sound fields. We will also purchase (with the help of capital equipment money from another request) and test a programmable computer-controlled amplifier and data acquisition system. We propose to test and evaluate the hardware, develop signal processing algorithms, and establish procedures for testing various types of materials.

Task B, Real-Time Processing. This is a technology transfer effort designed to convert our successful processing algorithms (which are computationally intensive) to a near real-time or interactive form for everyday use in programmatic applications. This will use special real-time signal processing hardware along with efficient algorithms.

Outside Contacts

We have interacted with three consultants regarding this work.

- B. D. Cook (University of Houston). Professor Cook has visited LLNL several times. Discussions pertained to system-identi-

fication, extrapolation, the AMS problem, layer problems, and pulse preprocessing. Cook also performed experiments and processed data. He has written several reports and papers [R1,2,5,6,9–13,15,16].

- L. Adler (Ohio State University). Professor Adler visited LLNL twice. Discussions pertained to plans for system-identification work and methods for layer-problem analysis. Adler and D. K. Lewis have produced four reports [R17–20].
- R. Weglein (an independent consultant). Weglein has consulted with us for the last three years. This year he worked with D. K. Lewis and S. E. Benson in the development of AMS techniques. He has written two reports for this project [R4,21].

Financial Status

Total expenditures on this project during FY 86 were $230 000.

Reports/Publications Fully or Partially Supported by this Project

R1. B. D. Cook, "Review of the Goals of Modeling the Acoustic Microscope," *University of* Houston report, September 30, 1985.

R2. B. D. Cook and J. T. Miller, "Measurement of Reflection of Plane Waves for Planar Liquid-Solid Interfaces Acousto-Optically," *Review of Progress in Quantitative NDE* (University of California, San Diego, August 3–8, 1986).

R3. R. D. Streit and G. A. Clark, "Research in Quantitative Ultrasonic Nondestructive Evaluation," *Engineering Research Annual Report FY 85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985), pp. 111–124.

R4. R. D. Weglein, "Materials Characterization of a Layered Structure," written report to G. A. Clark, Lawrence Livermore National Laboratory, Livermore, CA (September 6, 1986).

R5. G. A. Clark, D. M. Tilly, and B. D. Cook, "Signal /Image Restoration for Quantitative Nondestructive Evaluation," *NDT International* 19(3), 169–176 (1986).

R6. G. A. Clark, J. D. Brase, B. D. Cook, and D. M. Tilly, "A- and B-Scan Resolution Enhancement Using System Identification Techniques," *Review of Progress in Quantitative NDE* (Williamsburg, VA, June 23–28, 1985).

R7. G. A. Clark and F. L. Barnes, "Model-Based Parameter Estimation for Layer Problems in QNDE," *Review of Progress in Quantitative NDE* (University of California, San Diego, August 3–8, 1986).

R8. G. A. Clark, F. L. Barnes, and G. Thomas, "Digital Signal and Image Restoration for Weld Underbead Profiling," *Review of Progress in Quantitative NDE* (University of California, San Diego, August 3–8, 1986).

R9. G. A. Clark, D. M. Tilly, B. D. Cook, and J. M. Brase, "Bandlimited Signal Extrapolation for A- and B-Scan Resolution Enhancement," *Review of Progress in Quantitative NDE* (Williamsburg, VA, June 23–28, 1985).

R10. G. A. Clark B. D. Cook, R. L. McKinney, and G. D. Poe, "Impulse Response of a Crack: Case Study of Signal Processing of Ultrasonic Signals," *Review of Progress in Quantitative NDE* (University of California, San Diego, July 8–13, 1984).

R11. E. K. Miller, G. A. Clark, G. D. Poe, D. D. Cook, and J. A. Jackson, "Combining Pole and Ramp-Based Techniques for Target Identification," *Proc. Advisory Group for Aerospace Research and Development, 1984 (35th Symposium, London,* 1984). See also Lawrence Livermore National Laboratory, Livermore, CA, UCRL-90363 (1984).

R12. B. D. Cook, G. A. Clark, R. L. McKinney, and G. D. Poe, "Impulse Response of a Crack: Case Study of Artifacts Produced by Transducer Misalignment," *Review of Progress in Quantitative NDE* (University of California, San Diego, July 8–13, 1984).

R13. B. D. Cook, G. A. Clark, and G. D. Poe, "Application of Signal Extrapolation to Ramp Response Inverse Problems," *Review of Progress in Quantitative NDE* (University of California, San Diego, July 8–13, 1984).

R14. R. D. Streit and G. A. Clark, "Research in Quantitative Ultrasonic Nondestructive Evaluation," *Engineering Research Annual Report FY 85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985), pp. 11–124.

R15. B. D. Cook, "A Real-Time Ultrasonic Flaw Sizing Scheme," *Review of Progress in Quantitative NDE* (University of California, San Diego, August 3–8, 1986).

R16. B. D. Cook and G. D. Poe, *Lommel Diffraction Correction for Different Size Transducers: Gaussian-Laguerre Formulation*, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-92376 (1985).

R17. S. I. Rokhlin, D. K. Lewis, K F. Graff, and L. Adler, "Real-Time Study of Frequency Dependence of Attenuation and Velocity of Ultrasonic Waves During the Curing Reaction of Epoxy Resin," *J. Acoustic Soc. Am.* 79, 1786 (1986).

R18. D. K. Lewis, "Characterization of Flaws and Structural Integrity by Ultrasonic Spectroscopy," Ph.D. thesis, Ohio State University, Columbus, OH (1986).

R19. L. Adler, "Ultrasonic Methods to Evaluate Surface Roughness, Theoretical Summary," Ohio State University, letter report to Lawrence Livermore National Laboratory, Livermore, CA (February 1985).

R20. D. K. Lewis, "The Effects of Rough Surfaces on the Ultrasonic Characterization of Defects," letter report to Lawrence Livermore National Laboratory, Livermore, CA (December 1985).

R21. R. D. Weglein and A. K. Mal, "Toward Quantitative Imaging with the Acoustic Microscope," to be published.

R22. T. G. Johnson, "Poe-Sherman Algorithm for Estimation of the Impulse Response of a Linear Time-Invariant System" (University of California, Davis, SIPL-85-13, July 1985).

R23. G. A. Clark, M.A A. Soderstrand, and T. Johnson, "Transform Domain Adaptive Filtering Using a Recursive DFT," *IEEE International Conference on Circuits and Systems* (Kyoto, Japan, June 5–7, 1985).

1. R. D. Streit and G. A. Clark, "Research in Quantitative Nondestructive Evaluation," *Engineering Research Annual Report FY 85*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-19323-85-2 (1985), pp. 111–124.
2. A. Atalar, "An Ansular-Spectrum Approach to Contrast in Reflection Acoustic Microscopy," *J. Appl. Phys.* **49**, 5130–5139 (1978).
3. J. R. Sheppard and T. Wilson, "Effects of High Angles of Convergence on $V(z)$ in the Scanning Acoustic Microscope," *Appl. Phys. Lett.* **38**(1), 858–859 (1981).
4. L. Ljung and K. Glover, "Frequency Domain Vs. Time Domain Methods in System Identification," *Automatica* **17**(1), 71–86 (1981).
5. D. M. Goodman, "NLS: A System Identification Package for Transient Signals," Lawrence Livermore National Laboratory, Livermore, CA, UCID-19767 (1983).
6. G. A. Clark, L. C. Martin, and E. J. Bogdan, "Identification of Antenna Parameters from Time-Domain Pulse Response Data," Lawrence Livermore National Laboratory, Livermore, CA, UCID-19770 (1983).
7. G. C. Goodwin and R. L. Payne, *Dynamic System Identification, Experiment, Design and Data Analysis* (Academic Press, New York, 1977).
8. L. Ljung, "Aspects on the System Identification Problem," *Signal Processing* **4**, 445–460 (1982).
9. P. Eykoff, System Identification (Wiley & Sons, New York, 1974).
10. J. Bendat and A. Piersol, *Engineering Applications of Correlation and Spectral Analysis* (Wiley-Interscience, 1980).
11. J. V. Candy, G. A. Clark, D. M. Goodman, and D. B. Harris, "Transient Electromagnetic Signal Processing: An Overview of Techniques," *Time Domain Measurements in Electromagnetics*, Michael Gardner, Ed. (D. Van Nostrand Co., Inc., Princeton, NJ), in preparation.
12. A. Papoulis and C. Chamzas, "Improvement of Range Resolution by Spectral Extrapolation," *Ultrasonic Imaging, An International Journal* **1**, 121 (1979).
13. A. Papoulis, "A New Algorithm in Spectral Analysis and Band-limited Extrapolation," *IEEE Trans. Circuits and Systems* **CAS-22**(9) (1975).
14. J. A. Cadzow, "An Extrapolation Procedure for Band-Limited Signals," *IEEE Trans. Acoust. Speech and Sig. Proc.* **ASSP-27**(1) (1979).
15. R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained Iterative Restoration Algorithms," *Proc. IEEE* **69**(4) (1981).
16. D. C. Youla, "Generalized Image Restoration by the Method of Alternating Orthogonal Projections," *IEEE Trans. Circuits and Systems* **CAS-25**(9) (1978).
17. J. M. Mendel, *Optimal Seismic Deconvolution, an Estimation-Based Approach* (Academic Press, 1983).
18. J. M. Mendel and J. Goutsias, "One-Dimensional Normal-Incidence Inversion: A Solution Procedure for Band-Limited and Noisy Data," *Proc. IEEE* **74**(3), 401 (1986).
19. E. Habibi-Ashrafi, "Estimation of Parameters in Lossless Layered Media Systems," *IEEE Trans. Automatic Control* **AC-27**(1), 31–48 (1982).
20. J. M. Mendel, "Some Modeling Problems in Reflection Seismology," *IEEE ASSP Magazine* **4** (April 1986).
21. J. M. Mendel, "A Survey of Approaches to Solving Inverse Problems for Lossless Layered Media Systems," *IEEE Trans. Geoscience and Remote Sensing* **GE-18**(4), 320 (1980).

Development of Centrifuge Modeling for Evaluating the Mechanisms of Collapse of Underground Openings

B. C. Davis
Nuclear Test Engineering Division

L. Chang
Department of Civil Engineering
University of California, Davis

B. L. Kutter
Department of Civil Engineering
University of California, Davis

Improved prediction of surface collapse above an underground cavity is important in many LLNL programs, including nuclear testing. To improve our predictive capability it is necessary to achieve a better understanding of the mechanisms involved in the process of collapse. Our research aims to develop the centrifuge technique for modeling mechanisms of underground collapse in soil. We will also evaluate the adequacy of existing constitutive models of soil in relation to the problem of underground collapse.

During FY 86, using the centrifuge at U.C. Davis, we developed the basic centrifugal modeling technique and conducted tests, which included analytical modeling with the code NIKE2D. Our initial results indicate that the effects of various important parameters can be identified. We will work toward quantification of these effects in FY 87 and will continue modeling with NIKE2D.

Introduction

The objective of this research is to develop the centrifuge technique for modeling mechanisms of collapse in undergound openings in soil, and also to evaluate the adequacy of existing constitutive models of soil for use in modeling this type of collapse.

When an opening is excavated in a geologic medium, the load formerly carried by the excavated material will be shared by media surrounding the opening and support forces within the opening. If strains within the elastic limit are generated, this redistribution of stress can be readily approximated with elastic theory, physical models, or numerical techniques. However, the geometry of stress redistribution and the resulting deformations are difficult to calculate when large, nonlinear strains are involved because of the complex interaction of the various mechanisms that come into play. As a consequence, we cannot now adequately predict what happens when collapse occurs.

In connection with a quasi-static formulation of the underground collapse problem, we are currently (FY 86) investigating various model parameters, including the strength parameters of the surrounding geologic media, the distance of the opening below ground surface, and the effect of gravity on these parameters, as well as the boundary effect of the scale modeling. In FY 87 we will expand our research by increasing the sophistication of our model. The new, more complex model will incorporate additional parameters such as changes in opening geometry, added geological features, changes in initial stress field, and added complexity in the opening's surface. We will also identify an appropriate constitutive soil model for use with NIKE2D, a finite element code used to analyze the stress response of two-dimensional axisymmetric solids.

In FY 86 we initiated the centrifuge scale-modeling research using the Schaevitz Centrifuge (see Table 1) at the University of California, Davis, with joint funding by the Engineering Research Program and the Nu-

clear Test Program. In FY 87 we will continue to use the Schaevitz Centrifuge, and also plan to use the National Geotechnical Centrifuge, the largest machine of its kind in the world, which will allow evaluation of larger models.

Technical Status

Centrifuge Modeling

Scale model tests, if carefully conducted, can provide useful data when full-scale results are difficult or impossible to obtain. In scale model tests involving problems such as underground collapse where gravity is a major driving force, the forces of gravity must be scaled to make the model simulate the actual event or prototype. Gravity can be scaled by placing the models in a centrifuge, where they are subjected to a steady centripetal acceleration many times greater than the earth's gravity.

Table 1. Specifications of UC Davis centrifuges.

Radius to surface of platform	30 ft
Max. area of model	6 × 7 ft
Max. height of sample	5 ft
Max. payload mass	4 tons
Max. acceleration (estimated)	
1987	25 g
1988	100 g
1990	300 g
Radius to surface of platform	43 in.
Max. area of model	18 × 18 in.
Max. height of sample (approx)	16 in.
Max. capacity (sample weight × acceleration)	10,000 g-lb (e.g. 100 lb × 100 g)
Max. acceleration	175 g
Radius to surface of drum	2 ft
Circumference of model	12.6 ft
Width of model	10 in.
Max. height of sample	6 in.
Max. payload mass	600 lb
Max. acceleration	300 g

[a] Model is continuous around circumference of centrifuge.

We also consider other factors in our effort to maintain similarity between model and prototype. Mechanical properties such as permeability, shear strength, density, and moduli must be the same in the model as in the prototype. The stress fields in the model and prototype must also be the same, since the mechanical properties of the model are highly dependent on the confining pressures. For example, using the traditional Mohr-Coulomb equation:

$$\tau_f = c + \sigma_n \tan \phi \quad , \qquad (1)$$

the shear stress on a failure plane τ_f is defined as a function of two soil parameters, namely the cohesion (c) and the friction angle (ϕ), where the frictional contribution to the strength is proportional to σ_n, the normal stress on the failure plane.

Accurate modeling also requires that the mechanical properties of the model maintain an appropriate stress dependence to the prototype. As an example of stress dependence, particulate geologic media such as sand and soil tend to change volume (dilate or contract) during shear in a nonlinear and stress-dependent fashion. Thus the condition of identical stress is imposed:

$$\sigma_{model} / \sigma_{prototype} = \sigma^* = 1 \quad . \qquad (2)$$

This condition of identical stress leads to a set of scaling ratios (see Table 2), which must be maintained.

Limitations of Centrifuge Modeling

When applying the scaling laws to a particular problem, the researcher must keep in mind certain important limitations of centrifugal modeling. For example, the largest soil particle size must be much smaller than the critical, i.e., smallest model dimension. If this condition is not met, then the particle size must be scaled as all other dimensions are scaled, with special effort taken to maintain a similarity of the mechanical properties of the model soil (with adjusted grain size) to those of the prototype soil.

In addition, the centrifuge produces a nonuniform radial acceleration field proportional

to the radius of rotation. This means that the bottom of the model will be subjected to greater acceleration than the top. To minimize errors due to this nonuniform acceleration field, we keep the model dimensions much smaller than the centrifuge radius.

Relative motion within a model in a rotating reference frame can cause Coriolis accelerations. These accelerations are equal to

$$a_{cor} = 2\,\omega\,v_{rel} \quad , \qquad (3)$$

where ω is the angular velocity of the centrifuge. These forces are generally insignificant for quasi-static problems when v_{rel} is very small and also for most dynamic problems when the model accelerations or the model gravity factor is much greater than the Coriolis accelerations.

These limitations, and other possible scaling or boundary effects, indicate that centrifugal modeling should be verified by a model study that validates the modeling laws for the particular problem being studied. The best check on scaling laws would be a comparison to full-scale test data, but this is usually not possible. An alternative method evaluates the scaling laws by performing the same test at different scale factors. The testing of two different scale models is called "modeling of models." If both models predict the same prototype behavior, then the scaling laws and test procedure are verified over the range of scale factors spanned by the models. This method was used to test model cavities of 2-in. diameter and 6-in. depth at 38 g (N = 38), and cavities of 1-in. diameter and 6-in. depth at 76 g (N = 76). The 1-in. cavity was used as a model of the 2-in. cavity, and both represent a prototype cavity of 76-in. diameter and 228-in. depth. The test results validated the modeling laws as applied in the experiments reported here.

Experimental Program

Our tests begin with a container (the so-called model bucket or sample bucket) holding sand or alluvium soil, whose relevant properties, such as particle size, are carefully controlled. A schematic view of the test arrangement is shown in Fig. 1. The underground opening is simulated by a model cavity (MC), which is a rubber bag filled with water and buried in the sand or soil at the appropriate depth. To balance the soil overburden pressure on the MC and prevent its premature collapse, an appropriate water level is set in the supply reservoir with valve SV1 open and valve SV2 closed. During tests the water level in the supply reservoir and

Table 2. Summary of gravity-scaling ratios.

For all events		For dynamic events	
	Model dimension		Model dimension
Parameter	Prototype dimension	Parameter	Prototype dimension
Stress	1	Time	1/N
Strain	1	Frequency	N
Length	1/N	Velocity	1
Area	$1/N^2$	Acceleration	N
Volume	$1/N^3$	Strain rate	N
Mass	$1/N^3$		
Density	1		
Force	$1/N^2$		
For self weight		**For diffusion events**	
Time	Not applicable	Time	$1/N^2$
Acceleration	N	Velocity	N^2
		Acceleration	N^3
		Strain rate	N^2

Note: This table assumes that the model and prototype are constructed of material with identical mechanical properties; that the rate of strain does not affect the mechanical properties; and that no two important scaling ratios are conflicting.

the pressure in the MC are monitored by pressure transducers. With this arrangement the increase in overburden pressure as the centrifuge comes up to speed is exactly balanced by the accompanying increase in hydrostatic pressure inside the MC. After the centrifuge reaches the desired acceleration, SV1 is closed. To initiate the collapse of the MC, SV2 is opened for short intervals, letting water drain from the MC into the collection reservoir, whose volume is monitored by another pressure transducer. The flow restriction shown at A prevents oscillations in the system due to the operation of SV2. Figure 2 shows the test setup in place on the Schaevitz centrifuge at U.C. Davis, and Table 3 summarizes conditions and crater data for the 17 tests performed thus far.

Test Procedure

The tests were performed in Monterey 0/30 sand or Nevada Test Site (NTS) Area 2 alluvium. The NTS alluvium provided to us contained a significant amount of gravel and small boulders. We removed all particles larger than 0.2 in. by sieving to produce the grain size distribution labeled "coarse alluvium" in Fig. 3. This soil was used for most of the tests requiring a cavity of 2-in. diameter. For tests with a cavity of 1-in. diameter the alluvium was further sieved, removing all particles larger than 0.11 in., to produce the grain size distribution labeled "fine alluvium" in Fig. 3, which also includes the grain size distribution for the Monterey sand. Table 4 summarizes some representative triaxial test results for these soils.

Before the soil was placed in the model bucket, the rubber bag was installed near the Plexiglas window. A pluviator was then used to place the Monterey sand at a uniform density. Due to the gravel content of the NTS alluvium, a hand-pouring method was used for this material to avoid possible damage to the pluviator. The alluvium was poured from a height of about 10 in. from the top of the model bucket. When the model bucket was

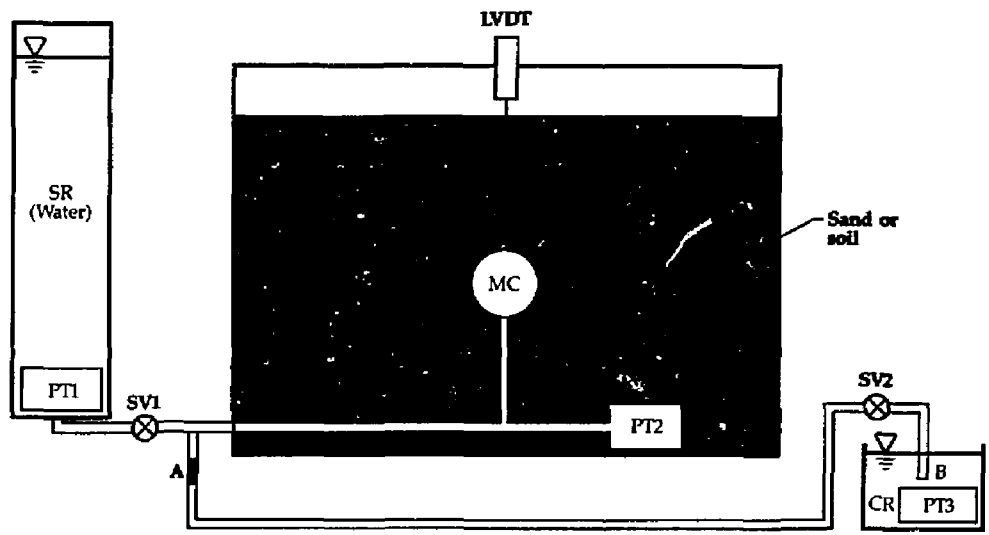


Figure 1. Schematic view of the test setup.

half full, it was vibrated at 65–68 Hz for 15 min and then the bucket was completely filled and vibrated again. In some tests (CC1–CC13, and also CC16–17 of Table 3) thin layers of an oil-based clay were used to indicate the deformation pattern. Colored alluvium was used for this purpose in tests CC14–CC15.

We recorded pressure changes throughout each test using pressure transducer PT2, which gives pressure in the model cavity (see Fig. 1). A typical test sequence is divided into five major segments, which are separated by significant periods of time, represented by breaks in the curve (see Fig. 4). Pressure data during these intervening

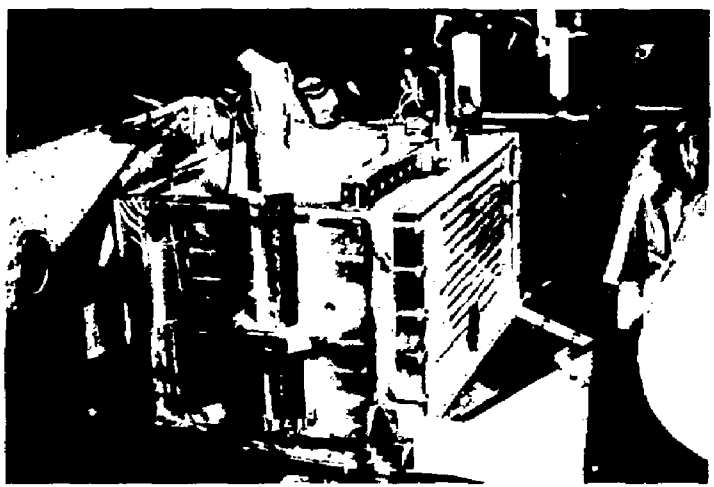


Figure 2. Test setup installed in the Schaevitz centrifuge.

Table 3. Summary of test conditions and crater data.

Test	Test description code[b]	Soil density (lb/ft³)	Centrifuge acceleration	Degree of saturation (%)	Crater volume (initial cavity volume)	Cavity volume change (initial cavity volume)
CC1	S3H2	109.7	76.5	0	—	—
CC2	S3H2	107.4	76.5	0	0.50	—
CC3	S3H2	111.9	76.5	0	0.45	0.78
CC4	S6H2	111.7	82.8	0	0.24	0.58
CC5	S6S2	110.6	82.8	0	0.44	0.80
CC6	C6H2	96.7	82.8	0	0.35	0.72
CC7	C6H2	101.0	82.8	0	0.52	0.67
CC8	F6H1	99.0	82.8	0	0	0.97
CC9	F3H1	100.0	76.5	0	0	0.93
CC10	F6H2	100.3	38.2	0	0.26	0.93
CC11	C6H2	100.0	38.2	0	0.12	0.87
CC12	F3H1	102.4	76.5	0	0.24	0.71
CC13	F3H1	102.4	76.5	0	0.38	0.67
CC14	C6H2	96.8	38.3	0	0	0.44
CC15	C6H2	98.9	38.3	0	0.19	0.58
CC16	C6H2	109.9	38.3	48[a]	0	0.81
CC17	C6H2	124.1	38.3	100	0	0.87

[a] This is an average degree of saturation. The water content actually increases with depth from about 8 to 14%.

[b] Test description code: e.g., S 3 H 2

- *Soil Type*: S = Monterey 0/30 Sand; C = Coarse Alluvium; F = Fine Alluvium
- *Cavity Depth in Inches*: 3
- *Cavity Shape and Location*: H = Hemispherically placed against window; S = Spherical in middle of sample.
- *Cavity Diameter in Inches*: 2

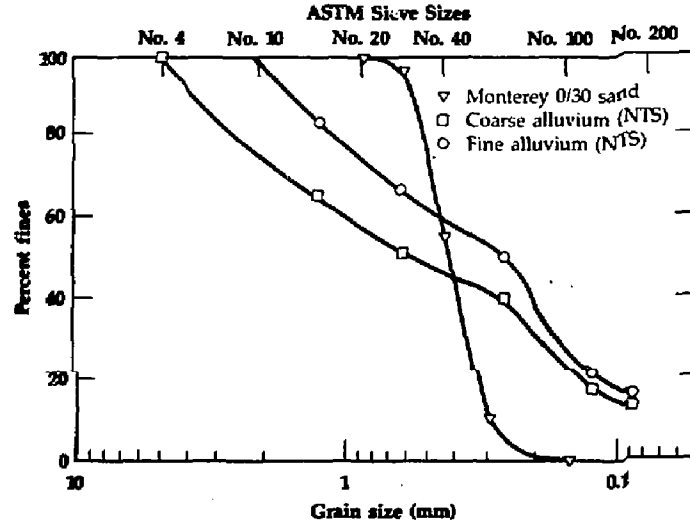


Figure 3. Grain size distribution of experimental soils.

Table 4. Summary of triaxial test results.

Soil	Density γ (lb/ft³)	Confining pressure σ_3' (psi)	Peak friction angle ϕ (degrees)	Young's modulus E' (psi)
Fine alluvium	90	14.0	42	1790
	90	7.2	44	2740
Monterey 0/30	106	3.0	52.4	
	106	7.7	48.1	
	106	14.5	45.92	
	106	29.0	44.9	
	106	58.0	43.4	
	106	100.0	41.9	

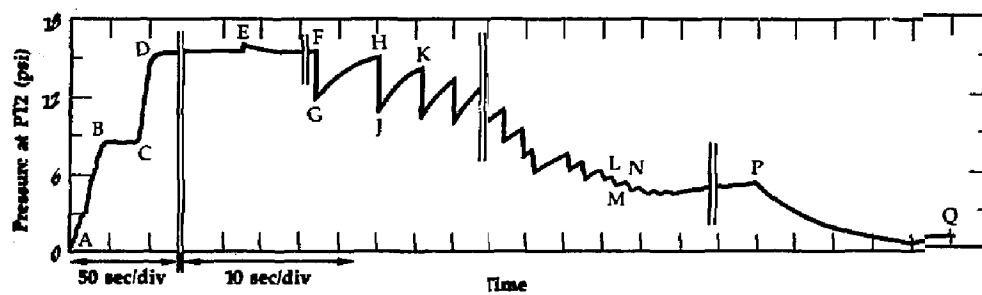


Figure 4. Typical test sequence as recorded at pressure transducer PT2.

periods are not relevant to the tests and are not shown.

At A in the test sequence (see Fig. 4) the centrifuge is at rest, and PT2 shows only a small pressure due to earth's gravity acting on the column of water in the supply reservoir. As the centrifuge speeds up to 21 g (136 rpm) the pressure increases. This speed is maintained from B to C and then increases to 38 g (185 rpm) at D. Here, at the first break in the curve, the time scale changes from 50 to 10 s/div. Then valve SV1 closes, producing a small pulse of pressure at E. After another interval, F marks the beginning of the opening and closing sequence for valve SV2. Each steep segment of the pressure curve (e.g., FG, HJ) shows when SV2 was briefly opened for less than 1 s, allowing a small amount of water to drain from the MC. The valve is then closed, and the pressure gradually increases until the next time SV2 is opened. This increase in pressure is caused by a subsidence of the soil around the MC, which reapplies soil pressure to the cavity.

At the beginning of the test sequence (FG) the soil surrounding the MC has only a moderate stress and is consequently quite stiff, taking a relatively long time to deform and transfer soil pressure to the cavity. However, as the opening-closing cycle continues to LM, soil pressures approach the collapse point. Now the soil is in a state of failure and no longer holds its shape, transferring soil pressure immediately to the cavity. This state is reflected in the pressure curve, which shows only a small pressure decrease as valve SV2 is opened. The experiment ends when no further volume change is recorded in the collecting reservoir, indicating the final stage of collapse. At this point (P, Fig. 4) the centrifuge is turned off and coasts to a stop (Q).

Mechanisms of Collapse

One advantage of centrifuge testing is the easy observation of deformation mechanisms. We observed two major modes of collapse, each of which showed slight variations. Nearly all the tests exhibited collapse by Mode 1 [Figs. 5(a) and 6(a)]. As the test progressed and cavity pressure decreased, the zone of soil affected by the collapse gradually increased, causing a decrease in cavity volume until stage D is reached, representing surface cratering. Some tests progressed only as far as stage B. Many parameters influence the ultimate development of the collapse, including soil type, soil grain size, moisture content, cavity depth and diameter, and differing boundary conditions.

In Fig. 6(a) the zone of soil affected by the collapse is bounded by solid vertical lines. This zone decreases in width (W) as the elevation above the cavity increases. At stage C [Fig. 6(a)] the effects of the collapse reach the surface. At this point the width of the affected zone starts to increase and a surface crater begins to form. We found that the width of the affected zone was influenced by depth of cavity, soil grain size, soil density,

Figure 5. Types of deformation observed in (a) tests CC1 through CC15 and (b) tests CC16 and CC17.

and moisture content. The zone of affected soil was narrowest with coarse grained alluvium, high density, and high centrifuge acceleration. A quantitative treatment of these effects requires further study.

The Plexiglas viewing window may affect the deformation patterns we observed. Friction may reduce the settlement close to the window [View D'-D', Fig. 6(a)], so displacements measured at the window may not be the actual maximum displacements. This is substantiated by the behavior of the rubber bag, which does not seem to have completely collapsed when viewed through the window. However, excavation after the test shows that the bag does totally collapse, but always folds back on itself [View D'-D', Fig. 6(a)].

A different failure mode was seen in a test with partially saturated soil [Figs. 5(b) and 6(b)]. In these tests soil collapse was limited and a new, arch-shaped cavity was formed. The free arch around the new cavity is characteristic of frictional material with some degree of cohesion, which in this case was apparently due to moisture accompanied by tensional stresses in the pore fluid. In the case of partially saturated soil above a water table, the total stress on the soil grains is given by:

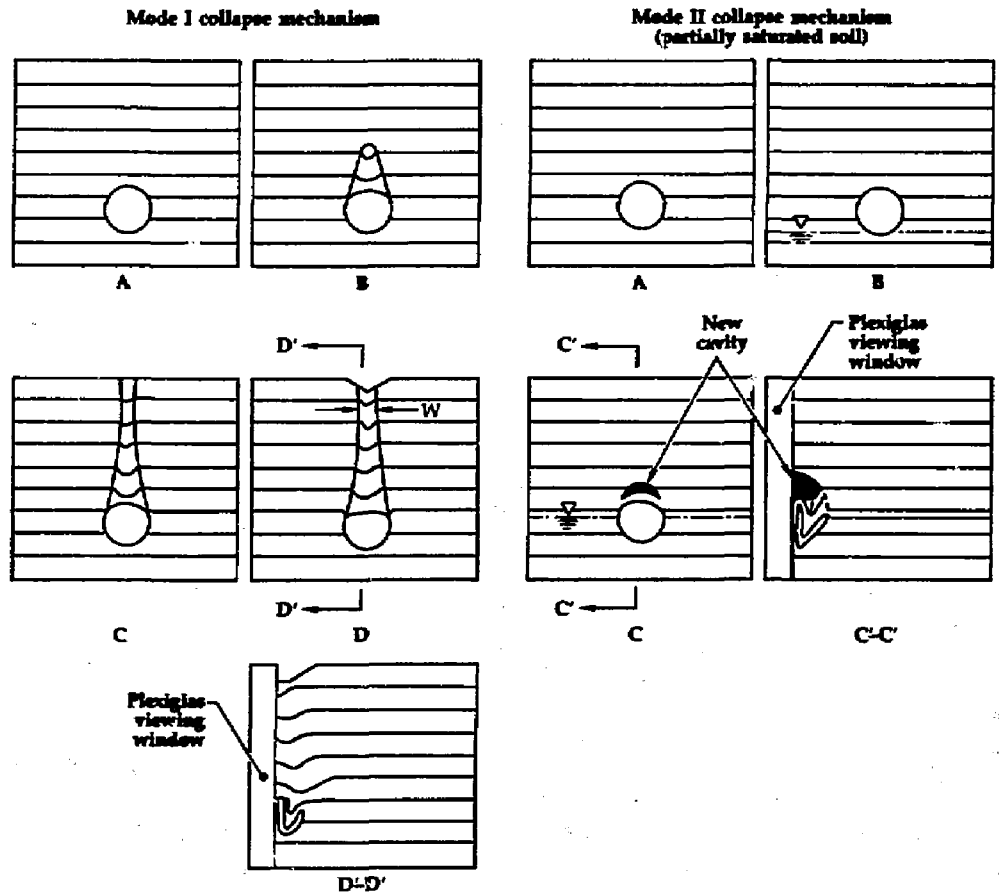


Figure 6. Mode of collapse observed in (a. 'ests CC1 through CC15 and (b) tests CC16 and CC17.

$$\sigma = \sigma' + u \quad , \tag{4}$$

where σ' is the effective normal stress on the soil grains, and u is the pore water pressure. For the above condition, if compressive stresses are taken to be positive, then the pore water pressure is negative. Thus even with zero total stress, there is a positive effective stress, which can give rise to frictional shear strength.

For a frictional material, the shear strength is given by the Mohr-Coulomb equation

$$s = \sigma'_n \tan \phi \quad , \tag{5}$$

which allows for a positive shear strength (s) if the effective normal stress (σ'_n) is positive. As indicated in Eq. (5), this can occur when the pore pressure is negative even if the total normal stress (σ_n) is zero. This relationship is relevant to test CC16. In this test the cavity pressure was reduced, but no collapse occurred because the soil had sufficient apparent cohesion to support an arch. Then water was added to the sample bucket, raising the water table. As the water table rose the degree of saturation increased and the negative pressure decreased until the apparent cohesion was reduced to a critical level. At this point the arch failed and soil slumped into the cavity. Presumably this would lead to a migration of the arch collapse to the surface if the water table were raised sufficiently.

Test Results

Table 3 summarizes the crater volumes for all tests, normalized by the initial volumes of the cavities. We found that the cavities did not ever collapse totally. Volume of the cavities after the tests was determined by submerging the rubber bag and measuring the volume it displaced. Table 3 also gives the relationship between crater volume and cavity volume change for all tests (see also Fig. 7). We found that the crater volumes were always less than the changes in cavity volume. Presumably this result is due to dilation of the soil as it settled into the cavity.

The graphical representation of our results (Fig. 7) shows considerable scatter, which may result from at least two factors. There was some variability in the test system as it evolved during the first six tests, causing significant changes in experimental procedure and sample preparation. Another factor could be variations in soil parameters. How-

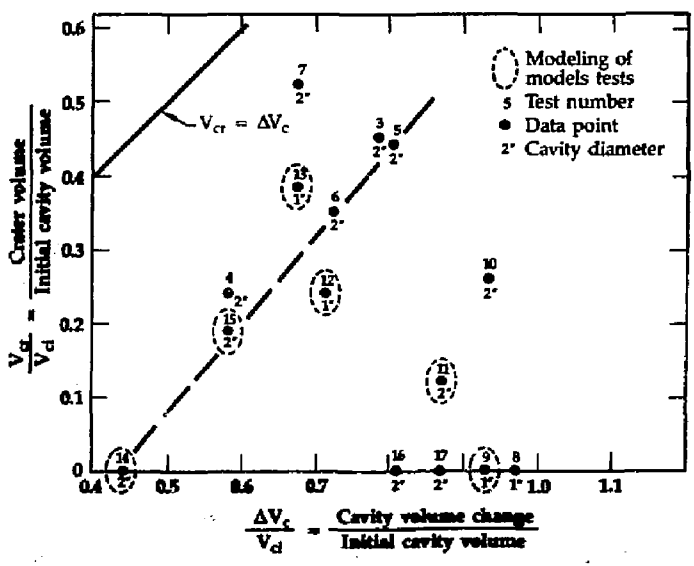


Figure 7. Crater volume vs change in cavity volume (normalized values).

ever, the last four "modeling of models" tests (CC12–CC15) appear to indicate that a significant relationship exists between cavity volume change and crater volume. Further testing is needed to confirm this relationship and the validity of the following general conclusions, which are based on our present crater and cavity volume measurements.

A number of tests were identical except for differences in soil density (CC6 and CC7; CC9, CC12, and CC13; CC14 and CC15). These tests showed that denser soil produced a larger crater. This result is contrary to intuition, which would suggest that a dense soil would dilate more than a loose soil and produce, under equivalent circumstances, a smaller crater. We further conclude that gravity appears to have a considerable influence on crater volume. At high centrifuge accelerations, crater volumes are larger (comparing test CC7 at 82.8 g with tests CC11, CC14, and CC15 at 38.3 g). Tests CC16 (moist soil) and CC17 (saturated soil) did not produce craters, probably due to water-induced soil cohesion as previously discussed.

Relationship of Cavity Pressure to Volume Change

We have compared normalized pressure in the model cavity to normalized change in cavity volume (Figs. 8 and 9). The normalized cavity volume is simply the change in cavity volume (ΔV) divided by the initial cavity volume (V). The normalized pressure, which we interpret as a measure of the driving force causing collapse, was ob-

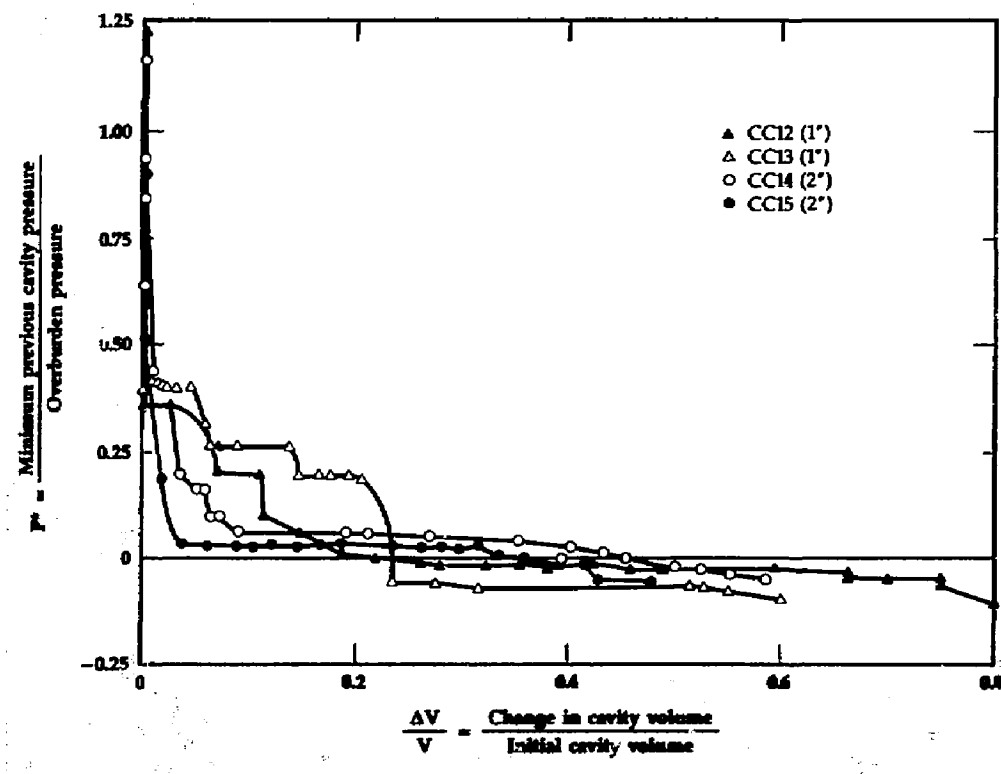


Figure 8. Minimum cavity pressure vs change in cavity volume (normalized) for modeling of models tests (CC12 through 15).

tained by dividing the minimum cavity pressure seen up to that moment in the experiment by the overburden pressure at the midpoint of the cavity, using the equation

$$P^* = [p_{PT2} - N\gamma_w h] / N \gamma z \quad , \qquad (6)$$

where p_{PT2} is the minimum cavity pressure recorded by transducer PT2 prior to collapse (see Fig. 1), N is the centrifuge acceleration, γ_w is the unit weight of water at 1 g, h is the difference in radius between PT2 and the center of the cavity, γ is the unit weight of the soil, and z is the depth from the surface to the center of the cavity. The minimum pressure reading was used because it was felt that the greatest proportion of cavity collapse seen at any point in the experiment would have been caused by the lowest previously observed value of internal cavity pressure.

The results of the final four tests conducted in dry alluvium are shown in Fig. 8. Tests CC12 and CC13 used fine alluvium, an acceleration of 76.5 g, and cavities of 1-in. diameter and 3-in. depth; tests CC14 and CC15 used coarse alluvium, an acceleration of 38.3 g, and cavities of 2-in. diameter and 6-in. depth. These tests, representing a "modeling of models" for a prototype cavity of 76.5-in. diameter and 230-in. depth, should give identical results, and in fact the normalized pressure histories as shown in Fig. 8 are very similar.

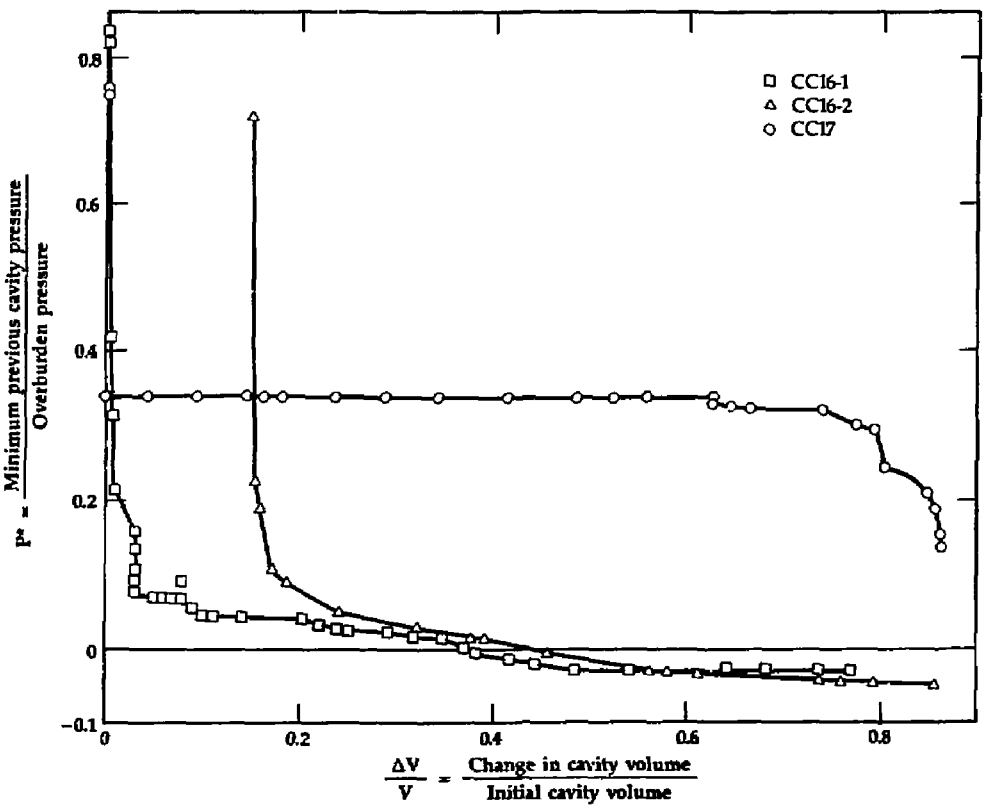


Figure 9. Minimum cavity pressure vs change in cavity volume (normalized) for partially saturated and saturated soil (CC16, CC17).

It is notable that visible signs of collapse begin when P* approaches 0.35 to 0.40. As P* approaches zero collapse continues, causing little change in pressure. Apparently a very small pressure is sufficient to support a cavity on the verge of collapse. The negative-trending values of P* near the end of the tests are an artifact of the calculations used to derive P*, since the actual pressure inside the bottom of the MC never falls below zero.

The pressure-volume relationships for tests in saturated and partially saturated soil are somewhat different (Fig. 9). In saturated soil, collapse occurred at a higher pressure, corresponding roughly to the pore water pressure.

Numerical Modeling

We compared calculations from code NIKE2D with our experimental results. The NIKE2D code, written by J. Hallquist, can model large elastic-plastic deformations, variable material properties, and sliding interfaces. Figure 10 shows an example of NIKE2D modeling on a cavity of 1-in. diameter at a depth of 3 in. The NIKE2D code does not yet model underground collapse accurately. It is limited by the resident elastic-plastic Von Mises model, which does not account for yield as a function of the confining stress. This problem is compounded by the reduction in normal stresses during collapse, which causes a strength reduction in frictional soil. The present version of NIKE2D also does not simulate the dilation or contraction of soil that occurs during the collapse process. We plan to add a bounding surface constitutive model to NIKE2D to accommodate these problems, which are unique to particulate materials.

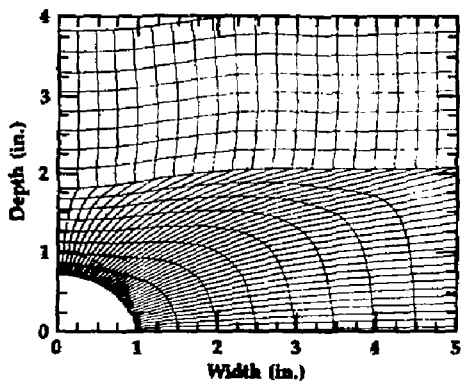


Figure 10. Example of NIKE2D modeling using von Mises criteria.

Conclusions

We have devised a method for centrifugal gravity-scale modeling of the collapse of an underground opening. Our experiments have validated our methodology, showing that results are replicable and that parametric studies can now be conducted to evaluate the influence of various relevant parameters on the collapse process. One of our objectives for FY 87 is to produce a quantitative definition of the contribution of various parameters to the dynamics of underground collapse. This will in turn permit more accurate numerical modeling.

Future Work

In the coming fiscal year (FY 87) we plan to run experiments that will verify certain general conclusions concerning the effects of important parameters. We will then be able to undertake further experiments, which will add geologic and geometric complexity to the model, including addition of geologic structure, changes in cavity and opening geometry, changes in the initial stress field, and variations in the strength of the cavity surface.

We also plan an evaluation of a constitutive model that will lead to enhancement of the existing code NIKE2D. We intend to move the centrifuge modeling to the National Geotechnical Centrifuge, which provides the capability of modeling much larger prototypes.

Outside Contacts

This project is a joint venture with the University of California, Davis, and benefits from the expertise of Dr. Bruce L. Kutter, an

assistant professor at that institution. The project has also benefitted from discussions with Dr. James A. Cheney of U.C. Davis and Lewis Thigpen of LLNL. In addition, the project has funded a graduate student, Lawrence Chang, who will use the research for his M.A. thesis.

Financial Status

Mechanical Engineering research funds spent on this project during FY 86 total $69 500. Additional funds from the Nuclear Test Program total $70 000.

Butkovich, T. R. (1976), *Cavities Produced by Underground Nuclear Explosions*, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-52097.

Kaliakin, V. N., Y. F. Dafalias, and J. A. Cheney (in progress), *Extension of the Bounding Surface Soil Plasticity Model to Account for High Pressure*, Lawrence Livermore National Laboratory, Livermore, CA.

Kutter, B. L., R. Lather, L. M. O'Leary, and P. Y. Thompson (1986), "Gravity Scaled Tests on Blast-Induced Soil-Structure Interaction," article submitted to ASCE Geotechnical Engineering Division.

Preece, D. S. and H. J. Sutherland (1986), *Physical and Numerical Simulation of Fluid-Filled Cavities in a Creeping Material*, Sandia National Laboratory, Albuquerque, NM, SAND86-0694.

Terzaghi, K. (1943), *Theoretical Soil Mechanics* (John Wiley & Sons, New York).

Thigpen, L. (1984), *On the Mechanics of Strata Collapse Above Underground Openings*, Lawrence Livermore National Laboratory, Livermore, CA, UCID-20152.

Thigpen, L. and J. T. Rambo (1985), *On the Formation of Sinks Above Underground Cavities in NTS Alluvium*, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-92419.

Isothermal Compressor/Gas-Bubble to Liquid Heat Exchanger (ICHX)

G. L. Johnson
Nuclear Test Engineering Division

M. A. Hoffman
Department of Mechanical Engineering
University of California, Davis

The cost of a fusion reactor would be reduced by replacing helium compressors and intercoolers with an isothermal liquid jet compressor and direct-contact heat exchanger (ICHX). The ICHX is fundamentally a water-jet ejector coupled to a water-circulating and heat-removal system. Analytical modeling and small-scale tests indicate that none of the ICHX configurations we investigated had efficiencies high enough to be useful in the proposed application.

Introduction

Thermal-conversion powerplant designs associated with fusion reactors tend to be very costly. For example, closed-cycle gas turbines using helium in a Brayton cycle have been suggested as part of a tandem-mirror fusion reactor. Such turbomachinery and the associated heat exchangers are expensive, and their efficiency depends on complex technology. We propose to replace the compressors and intercoolers with a simpler and cheaper, albeit less efficient, isothermal liquid jet compressor and direct-contact heat exchanger system (ICHX).

This concept was originally proposed by Dr. G. Logan. The first phase of the apparatus design and assembly for this research project was completed by W. Neef with the assistance of B. Meyers and J. Pitts.

The ICHX (Fig. 1) would replace the conventional components shown in the upper shaded area; it has fewer moving parts and thus fewer sealing problems than conventional turbomachinery. Its greater potential for modularity permits a simpler plant design. Initial cost and maintenance of a liquid jet compressor should be significantly reduced in comparison to conventional components. In addition, the liquid-liquid heat exchanger would be much smaller and less expensive than the gas-liquid intercoolers and precooler.

A liquid-jet compressor (LJC) couples a water-jet ejector to a water circulating system (Fig. 2). Helium gas from the helium turbines and regenerator is injected at the entrance to the ejector throat along with a jet of liquid water. The incoming gas is compressed by quasi-isothermal mixing with the liquid jet, which has a high kinetic energy.

The gas and water entering the compressor do not mix immediately. In the entrance zone the gas and water remain essentially separate, although a gas boundary layer develops on the wall, and interfacial shear forces act between the surface of the liquid jet and the gas annulus. Within the throat but prior to the mixing zone, wave disturbances on the jet may reach critical magnitude, causing the jet to shed droplets into the annular space. As the flow proceeds through the compressor, the liquid jet loses more and more mass and ultimately disperses.

In the mixing zone the droplets compress the gas by exchanging energy with it. Downstream from the mixing zone these droplets coalesce, and the liquid becomes a continuous phase containing finely dispersed gas bubbles, in contrast to the annular flow upstream where the gas is the continuous phase. The diffuser at the exit of the venturi converts much of the remaining kinetic energy of the bubbly mixture to a higher, static pressure. In our proposed application, the gas and liquid are then separated in the

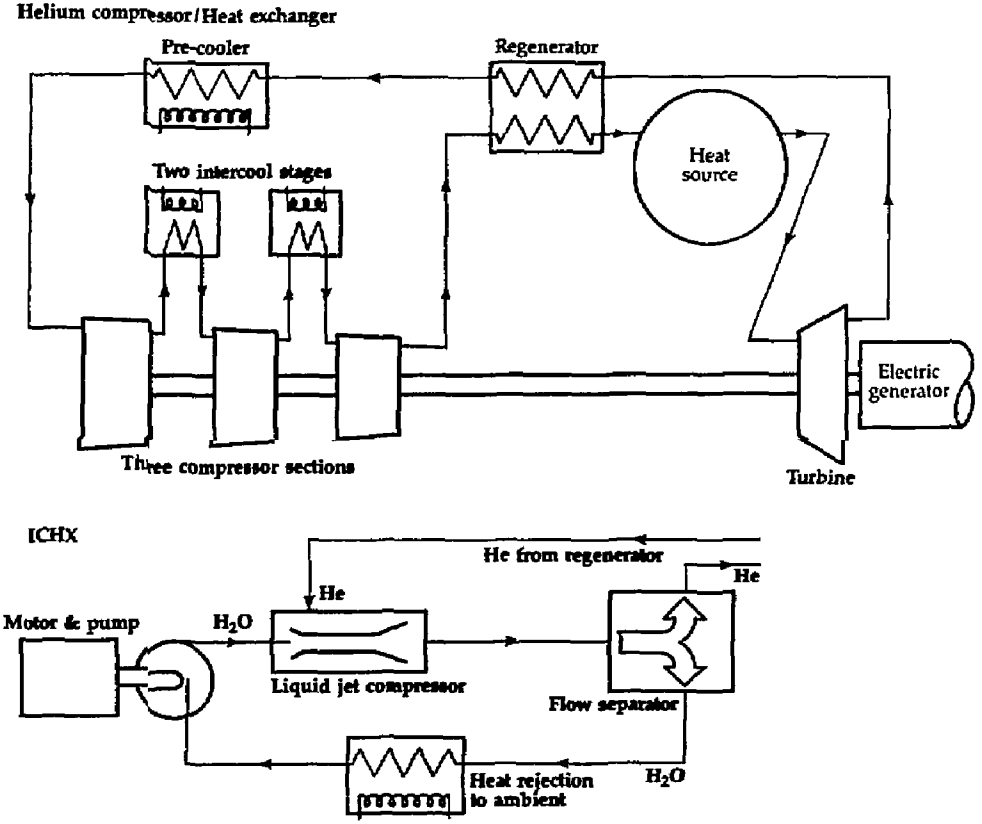


Figure 1. Comparison of ICHX with conventional compressor/intercooler.

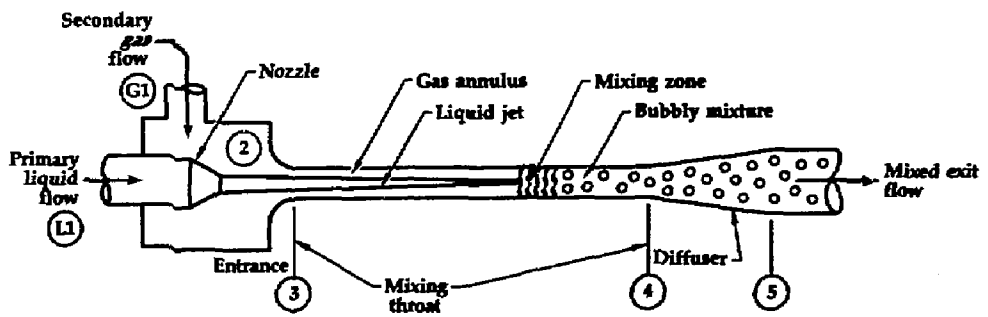


Figure 2. Schematic view of liquid-jet compressor (LJC).

ICHX system, and the compressed helium enters the regenerator (Fig. 1).

In FY 85 we reviewed this concept with Dr. R. G. Cunningham[1-3] and L. Hays, two major researchers in the field of liquid jet compressors. External contacts during FY 86 were primarily with M. Hoffman of UC Davis, a co-researcher on this project. Hoffman undertook most of the work on the analytical model development and provided extensive support in the test program and facility design.

The helium regenerative-gas-turbine cycle proposed for certain advanced fusion reactors has a maximum thermal efficiency with compressor pressure ratios ranging between 2 and 3, with the optimal ratio dependent on component efficiencies and other cycle parameters. Since nearly all of the LJC experimental results thus far published relate to higher pressure ratios, we developed a highly simplified model in FY 85 to estimate LJC efficiency at pressure ratios between 2 and 3, using loss coefficients published by Cunningham and Dopkin.[1-3] The results were very promising, indicating that LJC efficiencies of 75–80% were possible at a pressure ratio of 2. These figures imply an overall ICHX efficiency of 60–70%, assuming a motor and water pump efficiency of 80%.

We chose 65% as a working goal for ICHX efficiency. A value of 65% is lower than the efficiency of the equivalent intercooled compressor, which the ICHX is meant to replace; conventional gas-turbine compressor efficiencies are in the area of 80–85%. Lower ICHX efficiency results in lower cycle thermal efficiency (Fig. 3). However, associated reductions in projected capital and maintenance costs offset the lower ICHX performance, provided that the cost savings are about 15–20% of the total power plant capital costs.

At the beginning of FY 86 we developed a more complete quasi-one-dimensional model to refine our earlier performance estimates. Unfortunately, we found that certain terms in our equations that were neglected in the earlier, simplified model became very important at low pressure ratios. As a consequence, efficiencies calculated with our new model using Cunningham and Dopkin's relatively high loss coefficients were quite disappointing. The new model predicted ICHX efficiencies of less than 40% at pressure ratios of 2–3. It then became clear that the ICHX concept would be viable only if the loss coefficients were somewhat less than half as large as the values given by Cunningham and Dopkin.

These results caused us to revise our FY 86 program plan. We had originally planned an extensive test program including facility rework, injector redesign, and investigation of scaling effects. The pessimistic modeling predictions and a recommendation from the MER Midyear Review Committee resulted in a modification of this program. Our new test plan concentrated on determining actual device performance. For example, all tests to check the effect of scaling were entirely eliminated. We decided that unless testing revealed performance considerably better than our predictions, the test program would be terminated.

Technical Status

It has been shown[1-3] that static pressure changes across the LJC nozzle, throat, and diffuser can be described accurately by using quasi-one-dimensional continuity and momentum relations. This model requires

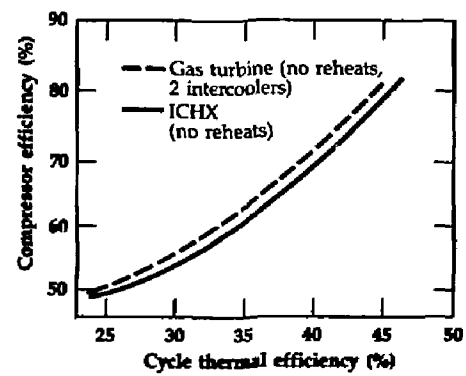


Figure 3. Cycle thermal efficiency vs required efficiencies for gas-intercooled compressor and ICHX liquid-jet compressor.

that all mixing between the two jets occur in the section of constant diameter called the mixing throat. We further assume that mixing is isothermal and is complete at the end of the mixing throat, and that no slip can occur between phases after mixing.

The following equations apply the continuity and momentum relations, including frictional forces. The pressure drop across the nozzle is given by:

$$P_{L1} - P_3 = Z_{L3}\,(1 + K_{nz} - B_n^4) \quad ,$$

where (referring to Fig. 2) P_{L1} is the pressure at the nozzle entrance, P_3 is the pressure of the liquid jet at the entrance to the mixing throat, Z_{L3} is the jet velocity head at location 3, K_{nz} is the nozzle component loss coefficient, and B_n^2 is the liquid flow area ratio A_{L3}/A_{L1}. The liquid velocity head Z_{L3} is based on an assumed vena contracta area $A_{L3} = C_3 \times A_n$ at the entrance to the mixing throat. The nozzle loss coefficient K_{nz} is dependent on the nozzle design as well as the Reynolds number.

The pressure drop across the gas chamber inlet, including the effect of compressibility, is given by a similar equation:

$$P_{G1} - P_3 = (D_R \times Z_{L3} \times \phi_0^2/c^2) \times (C_M + C_M \times K_{en} - B^4{}_G) \quad ,$$

where P_{G1} is the gas pressure at the inlet manifold, P_3 is the gas pressure at the entrance to the mixing throat, D_R is the density ratio ρ_{G3}/ρ_{L3}, Z_{L3} is the jet velocity head at location 3 (Fig. 2), ϕ_0 is the gas/liquid volume flow ratio, c is the gas/liquid area ratio at the mixing throat entrance, C_M is the compressibility coefficient, K_{en} is the component loss coefficient of the throat entrance, and B_G^2 is the gas flow area ratio A_{G3}/A_{G1}. The compressibility coefficient C_M is approximately equal to one plus a term proportional to the square of the gas Mach number at the throat entrance. This coefficient is usually set to one, since compressibility effects are negligible at our flowrates. K_{en} is usually equal to zero for our type of entrance region.

From a momentum balance on the constant diameter mixing throat, the pressure rise in the mixing throat is given by:

$$P_4 - P_3 = Z_{L3} \times [(2b') + (2D_R \times \phi_0^2/c^2) - (2 + K_{th}) \times (1 + D_R \times \phi_0) \times b'^2 \times (1 + P_3 \times \phi_0/P_4)] \quad ,$$

where P_4 is the pressure at the diffuser entrance, P_3 is the pressure at the entrance to the mixing throat, b' is the jet area ratio A_{L3}/A_3, and K_{th} is the component loss coefficient of the mixing throat. The loss coefficient K_{th} is an average value over the entire mixing throat. Solving the equation for the diffuser inlet pressure P_4 results in a quadratic equation. The solution to the quadratic equation uses the (+) term since we wish to solve for a pressure rise in the mixing throat.

The flow area, flow velocity, and density in the diffuser all vary with location. Assuming isothermal flow of a homogeneous two-phase mixture in the diffuser, the pressure rise in the diffuser is given by:

$$P_5 - P_4 = [Z_{L3} \times (1 + D_R \times \phi_0) \times b'^2] \times [(1 + P_3 \times \phi_0/P_4)^2 \times (1 - K_{di} \times b'^2) - a^2 \times (1 + P_3 \times \phi_0/P_5)^2] - [P_3 \times \phi_0 \times \ln(P_5/P_4)] \quad ,$$

where P_5 is the pressure at the diffuser exit, P_4 is the pressure at the diffuser entrance, and K_{di} is the component loss coefficient of the diffuser. The equation for the resulting diffuser exit pressure P_5 must be solved iteratively.

The LJC efficiency is defined as the useful pump power output for the isothermal compression divided by the liquid power input. This efficiency is given by:

$$\eta_{LJC} = P_{G3} \times \phi_0 \times \ln(P_5/P_{G2})/(P_{tot,L1} - P_5) \quad .$$

where $P_{tot,L1}$ is the total pressure of the liquid upstream of the nozzle. The pressure ratio across the LJC is given by:

$$PR_{LJC} = P_5 / P_{G2} \quad .$$

Note that the compressor efficiency goes to zero when the compressor pressure ratio

goes to one. Avoiding the region of sharp efficiency decrease is a problem when designing systems with low pressure ratios. The overall efficiency of an ICHX includes the efficiencies of the water pump and drive motor:

$$\eta_{ICHX} = \eta_{LJC} \times \eta_{H2OPUMP} \times \eta_{MOTOR} .$$

We generally assume the product of water pump efficiency and motor efficiency to lie between 80–90%.

Last year's simple model included rough approximations for many of these effects. A FORTRAN version of the new model was written for the IBM PC to calculate LJC pressure ratios and efficiencies as a function of LJC geometry and inlet flow conditions, as well as loss coefficients, which can be based on assumptions or experimental results.

The new FORTRAN code was used to generate a parametric analysis checking the effects of volumetric flow ratio and loss coefficient on LJC efficiency. A synopsis of the results is shown in Figs. 4 and 5. The analysis used the same liquid jet compressor geometry as the experiments described by Cunningham and Dopkins. The constant loss coefficients assumed for the first part of the parametric study are $K_{nz} = K_{en} = 0$, $K_{th} = 0.39$, and $K_{di} = 0.26$. These were calculated from the experimental conditions giving the best indicated efficiency. Loss coefficients were assumed to be constant over all parameter ranges, since their experiments did not completely cover our desired low pressure and flow ratios.

Cunningham and Dopkin found that increasing the flow ratio increases efficiency only up to a point. At a certain flow ratio, whose value depends on compressor geometry and jet velocity, the loss coefficients suddenly increase, causing a sharp decrease in efficiency. For this initial analysis we chose to be optimistic and keep the loss coefficients constant.

We had assumed that the minimum desired ICHX efficiency was 65%, based on the capital cost estimates for fusion power plant performance. With an assumed water pump/motor efficiency of 80%, the minimum LJC efficiency would be set at 82%. In the FY 85 MER study a simplified model was used, which suggested that these efficiencies could be reached with a reasonable amount of design manipulation. Using Cunningham and Dopkin's loss coefficients, the efficiency seemed to increase to near the desired performance as the pressure was decreased to the required ICHX operating range.

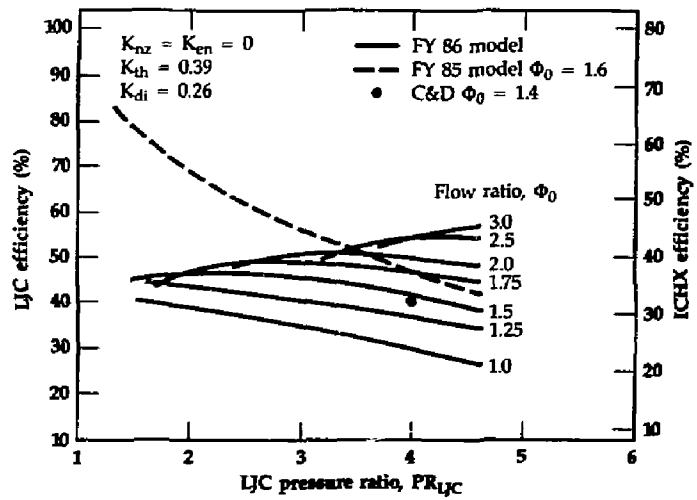


Figure 4. Comparison of FY 85 and FY 86 model predictions of LJC efficiency.

The FY 86 improved model compares well with data from Cunningham and Dopkin's results at a pressure ratio of 4 (Fig. 4). However, it predicts an efficiency far below that predicted by the simpler FY 85 model for the LJC compressor operating in the low pressure ratio range. The predicted efficiency is considerably below that needed by the ICHX even at the highest probable flow rate. In the desired low pressure range of 1.75 to 2.50, predicted LJC efficiencies are never greater than 50%. Operating at higher flow ratios does not give high efficiencies except at high pressure ratios. Even at the highest reasonable pressure and flow ratios the LJC efficiency is less than 60%.

We undertook further modeling to determine if efficiency would improve when the device was scaled up to reactor-sized pumps, flow conditions, and fluids. This analysis assumed the same loss coefficients as the small-scale model, and showed that for fixed values of flow ratio, nozzle/mixing throat area ratios, and pump number (pressure ratio-to-velocity head), the jet pump efficiency is almost independent of jet pump size, gas temperature, molecular weight, temperature level, and system pressure level.

Our analysis showed that efficiency could be improved only by reducing the loss coefficients. Figure 5 shows the effect of reduced loss coefficient on LJC performance. To attain a usable ICHX efficiency of 65% the loss coefficients would have to be lower than our most optimistic estimates (K_{th} = 0.2, K_{di} = 0.1).

It was plain that our concept was not feasible when modeled using the loss coefficients we had derived from published data. However, the Cunningham and Dopkin data from which the discouraging loss coefficients were derived was very limited in the pressure range of interest to us. We therefore believed that with a carefully designed experiment we might be able to demonstrate a compressor performance better than pre-

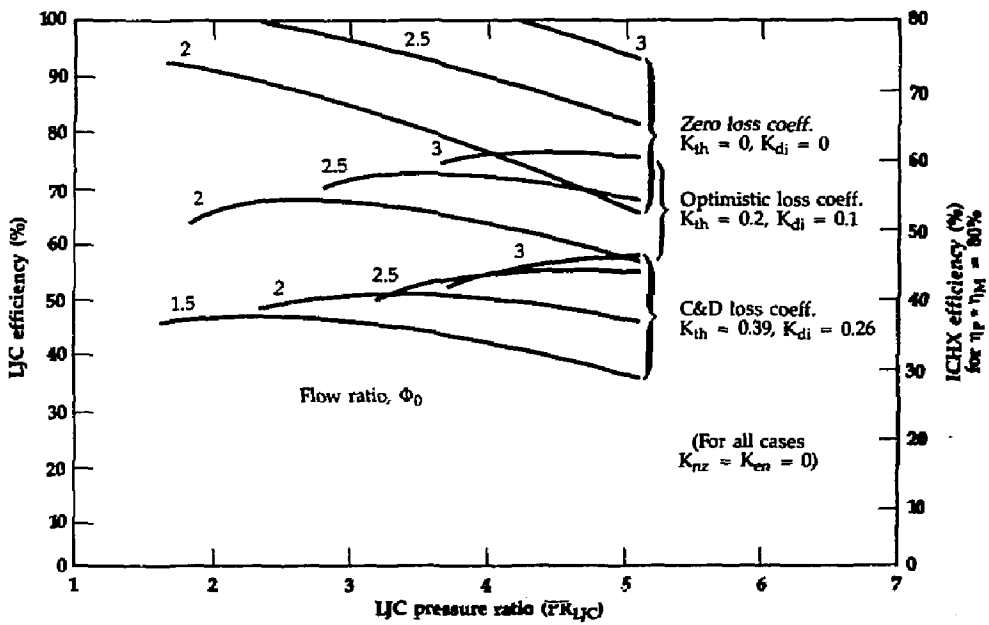


Figure 5. Effects of different loss coefficients on the efficiency of the liquid-jet compressor and the ICHX.

dicted, which would result in a substantial reduction in loss coefficients. Our entire experimental effort was focused on this goal.

Test Facility Improvements

At the end of FY 85 an LJC test facility was designed and built based on our experience at that time. Changes in this facility in FY 86 (Fig. 6) were based on FY 86 analytical work and a detailed review of requirements in the areas of data resolution, accuracy, and equipment control. Changes to the ICHX facility in FY 86 included:

- New pump added to give desired range of flow velocities and supply pressures.
- Water flow filter loop added to minimize particulate matter.
- Bypass loop added to control nozzle water flow rate.
- Primary water loop replumbed with flexible lines to provide for easy variation of the mixing throat length.
- Liquid-gas separator added to return line in reservoir.
- Flow control valves added to water loop and gas line for finer flow control.
- Nozzle design altered as suggested by Cunningham and Dopkin's findings.
- Filtered air supply added in conjunction with the N_2 gas supply to provide supplemental gas flow for cheaper, longer-term tests.
- Orifice meter added to primary water loop for more accurate water flow measurements.
- High-accuracy, high-resolution analog pressure gauges added, to meet required measurement accuracies.
- Pressure tap manifold added for measuring multiple pressure-tap locations with single gauge.

The pressure gauges and the gas flow rotameters were calibrated by the LLNL groups providing the instrumentation. The orifice meter was calibrated *in situ* using a stopwatch and weight scale to weigh the water flow collected over a given time.

Test Procedure

The experimental procedure involved setting the desired water flow and system pressures for the largest gas flow desired. After each data point was recorded the gas

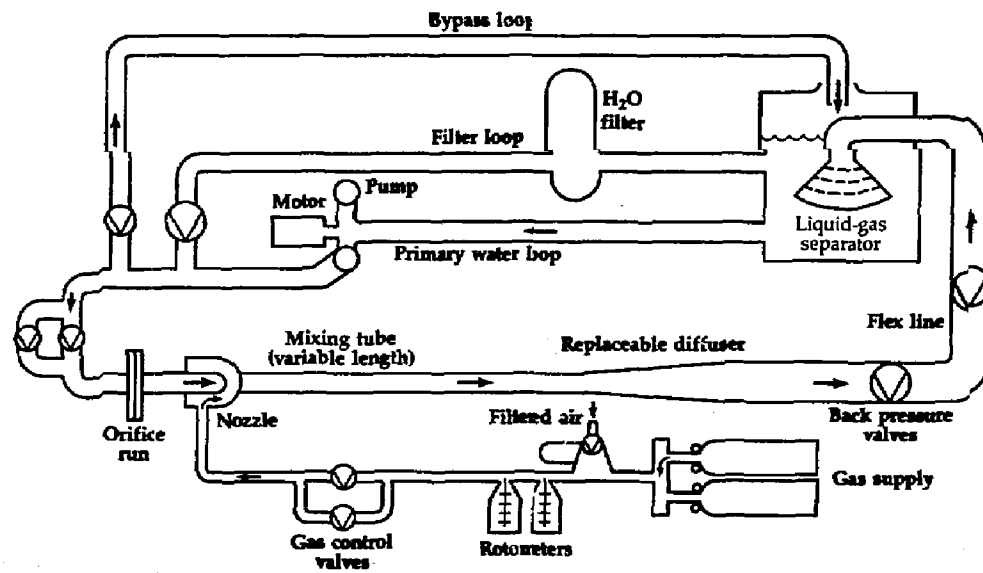


Figure 6. Upgraded experimental facility with changes made for FY 86 tests.

flow control valves at the LJC gas inlet and the back pressure valve at the LJC exit were throttled down to decrease the gas flow to a new flow rate while maintaining the gas inlet pressure. The water jet conditions remained constant with this procedure.

The system pressure ratio is obtained by dividing the measured diffuser exit pressure by the gas chamber pressure. The low system pressure ratios of interest to us are associated with a relatively low jet velocity head (Z_{L3}), i.e., a low water flow rate, or a relatively high gas chamber pressure (P_{G3}), i.e., a high system back pressure. The pressure ratio may be changed either by varying the gas inlet pressure or by varying the water inlet pressure. Our initial tests (Fig. 7) showed that ICHX low pressure ratio test results were essentially independent of the variable used. We chose to vary the gas chamber pressure.

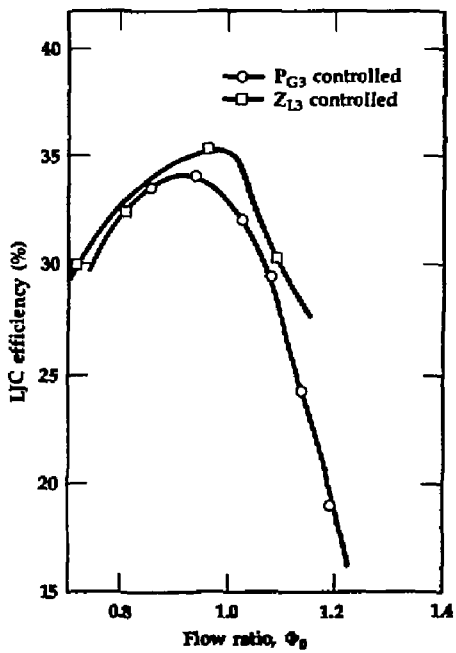


Figure 7. LJC low pressure ratio results, showing that results are independent of the variable used to control pressure ratio.

Test Results

To certify our ICHX facility results for low pressure ratio tests and multi-jet tests, we first tested an LJC test section at a pressure ratio of about four using the same geometric parameters as Cunningham and Dopkin's most efficient configuration (i.e., throat length L_T = 22, nozzle standoff distance S = 3, and jet area b' = 0.3). Our test data show values comparable to Cunningham and Dopkin for a pressure ratio of 4.0 as shown in Fig. 8. We felt that this adequately validated our facility. However, our measured LJC efficiency for pressure ratios close to 2.0 was much higher than Dopkin's. A check of Dopkin's thesis showed that he may have had some problems with his low pressure ratio performance tests.

Figure 9 shows maximum efficiency plotted against system pressure ratio. Although our facility could not generate measurements at Cunningham and Dopkin's high flow ratios, ϕ_0, required for maximum

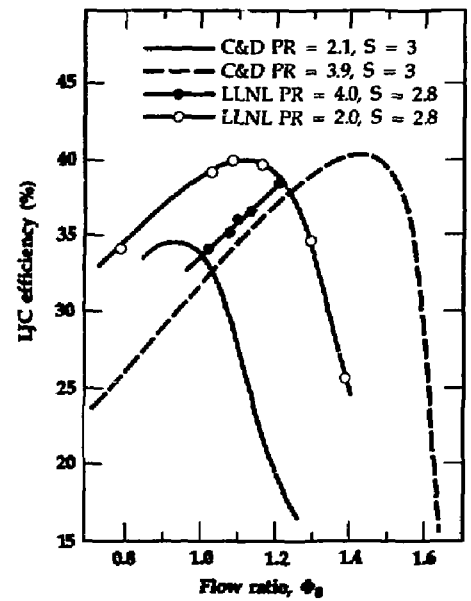


Figure 8. Comparison of some LLNL test results with Cunningham and Dopkin data.

efficiency at the higher pressure ratios, extrapolating our results to similar levels showed good agreement.

As shown in Fig. 9, the maximum efficiency seems to rise only slightly as the pressure ratio decreases. At a certain point the LJC efficiency begins to drop rapidly toward a value of zero. For a mixing throat length of 22.5 throat diameters (L_T = 22.5) this rapid drop occurs at a pressure ratio of approximately 2.0. When we tested a shorter mixing throat (L_T = 17.1) the sudden drop in efficiency occurs at a pressure ratio closer to 1.5 (Fig. 9). At low pressure ratios the maximum efficiency is slightly higher for the shorter mixing throat.

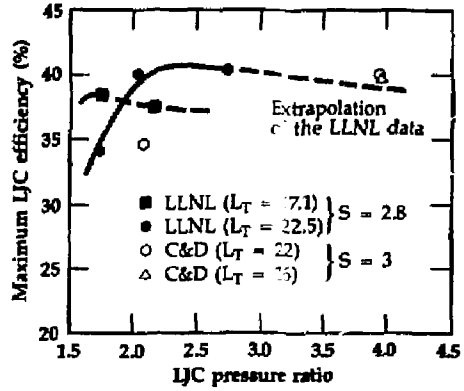


Figure 9. Maximum efficiency vs pressure ratio, showing agreement of LLNL data with previous results of Cunningham and Dopkin.

We attempted to gain some insight into the problem by investigating the pressure profiles inside the mixing throat. In the constant-diameter mixing throat section of the LJC, the liquid-gas mixing zone is characterized by a rise in the static pressure (Fig. 10). The mixing zone was approximately defined by the region of steepest slope of the pressure profile. The measured maximum efficiency of the LJC for medium to high pressure ratios always seemed to occur when the mixing zone ended just at the diffuser entrance, i.e., when the pressure gradient went to zero at that location. A pressure profile characteristic of maximum LJC efficiency is indicated in Fig. 10 by the curve with the black dots.

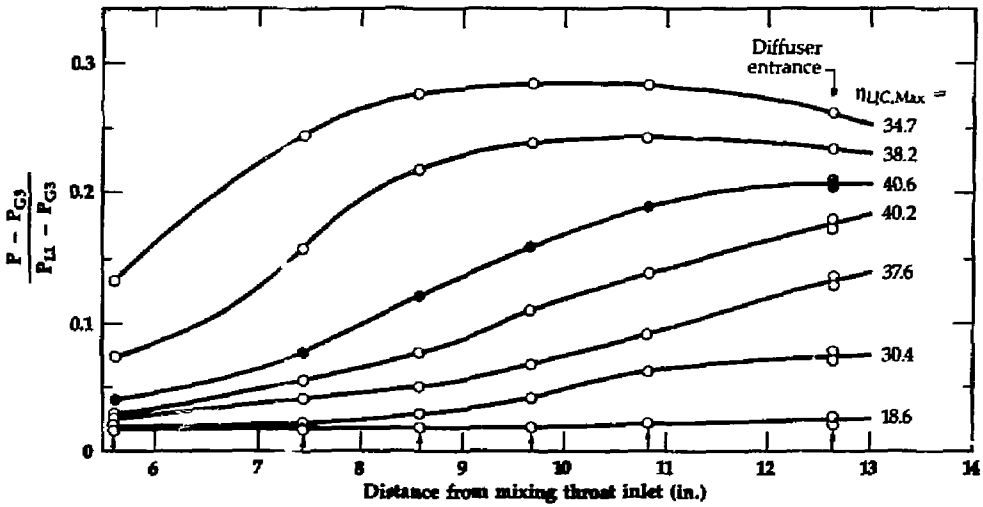


Figure 10. Pressure profiles in the mixing throat. Maximum efficiency is associated with a zero pressure gradient at the diffuser inlet, indicating that the mixing zone ends at that point (curve with black dots).

In the low pressure ratio experiments cited above, the mixing zone for the long mixing throat (L_T = 22) appeared to end upstream of the diffuser inlet. The LJC delivered higher maximum efficiencies when a shorter mixing throat (L_T = 17.1) was used with a length matching the mixing zone at that pressure ratio. However, none of our experiments at low pressure ratios yielded LJC efficiencies high enough to justify substitution of an LJC for an intercooled compressor in the proposed fusion reactor application.

We had hoped that multi-jet injection would improve LJC efficiency. We accordingly ran tests on a seven-hole nozzle similar to a design tested by Cunningham and Dopkin. The total flow area was the same as in the single-orifice nozzle, but was equally divided among seven orifices. We expected that multi-jet performance would be very sensitive to the standoff distances, i.e., the distance between the nozzle and the entrance to the mixing throat. As a check on the sensitivity of single-jet LJCs to this parameter, we compared our performance for two standoff distances with Cunningham and Dopkin's results. The variation in normalized maximum efficiency as a function of dimensionless standoff distance ratio S is shown in Fig. 11. Our LJC results fall on the same curve. Cunningham and Dopkin reference the standoff distance ratio S to the mixing throat diameter. We have also defined a modified standoff distance ratio S' referenced to the nozzle orifice diameter. We felt that comparisons between single jet and multi-jet performance should be based on the S' parameter, since the jet in the gas chamber (Station 2 in Fig. 2) acts more like a free jet.

One multi-jet test was run at the same standoff distance (S = 3, S' ≈ 10) as our single-jet test (S = 3, S' = 4). The modified standoff distance of a second multi-jet test (S' = 5) was nearly the same as the single-jet condition (S' = 4). The results of both tests at a pressure ratio of about 2.5 (Fig. 12) show a substantial decrease in the LJC efficiency for the seven-hole nozzle, although the S' = 5 test showed a 5 percent improvement in efficiency over the S' = 10 test.

Conclusions

We conclude that replacing a helium intercooled compressor with an ICHX is not feasible unless a dramatic improvement in performance is obtained, possibly through experiments involving equipment of large scale. Our results with the small-scale appa-

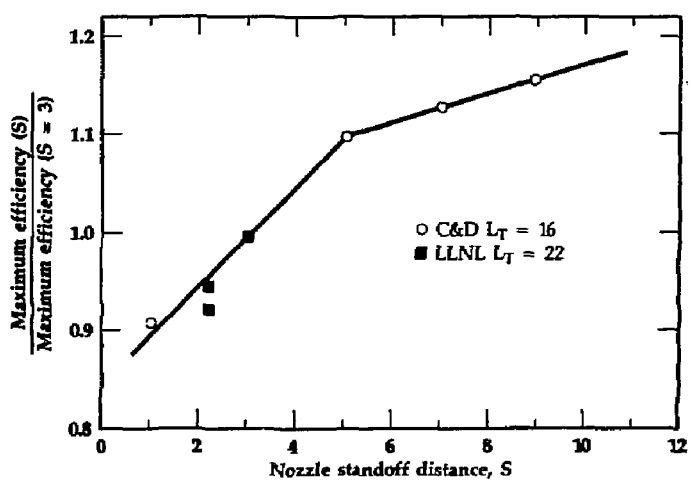


Figure 11. Efficiency as a function of standoff distance, showing agreement between LLNL data and results of Cunningham and Dopkin.

ratus were uniformly discouraging. We found that neither low pressure ratio operation nor multi-jet nozzle design gave any significant improvement in LJC performance as compared to the operation of a single-jet device at high system pressure ratios.

Future Work

No future work is recommended using the small-scale apparatus. If further research is undertaken, it should involve larger-scale equipment.

Financial Status

Engineering research funds spent on this project during FY 86 total $125 000.

1. R. J. Dopkin, *The Liquid-Jet Gas Pump: A Study of Jet Breakup and Required Throat Length*, M.S. Thesis, Pennsylvania State University, Pittsburg, PA (1973).
2. R. G. Cunningham, "Gas Compression With the Liquid Jet Pump," *J Fluid Eng.* 96, 203 (1974).
3. R. G. Cunningham and R. J. Dopkin, "Jet Breakup and Mixing Throat Lengths for the Liquid Jet Gas Pump," *J Fluid Eng.* 96, 216 (1974).

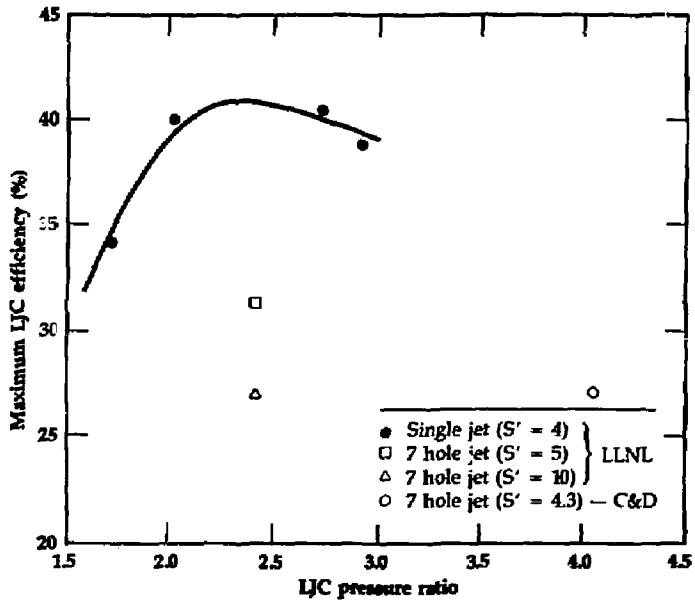


Figure 12. Efficiencies of single-jet and multi-jet nozzles, showing much better performance of single jets for this small-scale apparatus.

Author

Subject