Proceedings of the

FIFTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES

Instrumentation, Diagnostics, and Material Behavior

June 17-19, 1987

at

Argonne National Laboratory
Argonne, Illinois

Cosponsored by

Office of Basic Energy Sciences
U. S. DEPARTMENT OF ENERGY

and

Materials and Components Technology Division
ARGONNE NATIONAL LABORATORY

Coordinated by

Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed herein to any specific commercial product, process, or service. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
FOREWORD

The Fifth Symposium on Energy Engineering Sciences was held on June 17-19, 1987, at the Argonne National Laboratory, Argonne, Illinois. These proceedings include the program, list of participants, and the papers that were presented during the eight technical sessions held at this meeting.

This was the fifth Symposium sponsored by the Engineering Research Program of the Office of Basic Energy Sciences of the U. S. Department of Energy. As the title indicates, the central themes of this year's meeting were Instrumentation, Diagnostics, and Material Behavior. Each year a group of researchers in the DOE/BES Engineering Research Program are invited to present their research findings in such an open forum. By organizing around the three central topics principal investigators, their associates, and other interested parties were able to engage in discussions and share thoughts on subjects of common interest. This format has been used in the Engineering Research Program rather than annual contractors' meetings and has been extremely successful since its inception in 1983.

This Symposium was organized into eight technical sessions; three were titled Material Behavior I, II, III, respectively, and the others addressed the subjects of Plasma Processes, Fluid Mechanics/Robotics, Non-Destructive Evaluation, Welding Processes, and Non-Linear Processes. Papers in each of these subject areas approached the topics in a variety of ways, spanning the spectrum from pure theory to totally experimental studies. Contributions dealt with such diverse topics as fracture mechanics, non-imaging optics, optical tomography, and elastomeric laminates, among others. The breadth of topical material is indicative of the rich mixture of research supported by the DOE Engineering Research Program.

It is appropriate to restate here the goals and mission of the Engineering Research Program. The DOE Office of Basic Energy Sciences, of which Engineering Research is a component program, is responsible for the long-term mission-oriented research in the Department. It has the prime responsibility for establishing the basic scientific foundation upon which the Nation's future energy options will have to be identified, developed, and built; it is committed to the generation of new knowledge necessary for the solution of present and future problems of energy exploration, production, conversion, and utilization.

Consistent with the DOE/BES mission, the Engineering Research Program is charged with the identification, initiation, and management of fundamental research on broad, generic topics addressing energy-related engineering problems. Its stated goals are: 1) to extend the body of knowledge
underlying current engineering practice so as to create new options for enhancing energy savings and production, for prolonging useful equipment life, and for reducing costs without degrading industrial production and performance quality; and 2) to broaden the technical and conceptual base for solving future engineering problems in the energy technologies.

To achieve these goals, the Engineering Research Program supports over one hundred and ten research projects covering a wide spectrum of topics cutting across traditional engineering disciplines, with a focus on the following three areas: 1) mechanical sciences, 2) system sciences, and 3) engineering data and analysis. The Fifth Symposium on Energy Engineering Sciences covered the review of approximately one-third of the total activities sponsored by the DOE/BES Engineering Research Program.

The Fifth Symposium was held under the joint sponsorship of the DOE Office of Basic Energy Sciences and Argonne National Laboratory. The success of the meeting is directly attributable to the active participation of the researchers involved in the program. Several of the participants were also of great help as session chairpersons. Local arrangements for the Symposium have been handled by the Conferences Services Office of ANL under the able direction of Mrs. M. L. Holden. Mrs. Joyce A. Kopta of the ANL Technical Information Services Department assembled these proceedings and saw to their publication.

I am grateful to all who contributed to the success of the program, particularly to the participants who instilled an atmosphere of intellectual inquiry and excitement. I also wish to acknowledge the assistance of Dr. Lloyd Weber, for helping to arrange the symposium while working in the program office while on temporary assignment from the National Bureau of Standards.

James R. Welty, ER-15
Division of Engineering and Geosciences
Office of Basic Energy Sciences
FIFTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES
June 17-19, 1987
Argonne National Laboratory
Argonne, Illinois

FINAL PROGRAM

WEDNESDAY, JUNE 17

12:00 noon
REGISTRATION

12:45 p.m.
Opening remarks: The DOE Engineering Research Program
Oscar P. Manley, U.S. Department of Energy

12:55 p.m.
Introduction to the Fifth Symposium
Dr. James R. Welty, U.S. Department of Energy

TECHNICAL SESSION 1 - MATERIAL BEHAVIOR I
Chair: Tom Bruno, National Bureau of Standards

1:00 p.m.
"High Speed Optical Tomographic Data Acquisition Systems for Combustion Research"
L. Hesselink, Stanford University

1:30 p.m.
"Electron Attachment Phenomena in Heated Plasmas"
A. Chutjian, Jet Propulsion Laboratory

2:00 p.m.
"High Temperature Electronics"
T. Zipperian, Sandia National Laboratory

2:30 p.m.
"Semiautomated Burnett PVT Apparatus; Properties of a Geothermal Working Fluid"
J. M. H. Levelt Sengers, National Bureau of Standards

3:00 p.m.
"Properties of Elastomeric Laminates"
S. K. Clark, University of Michigan

3:30 p.m.
BREAK
TECHNICAL SESSION 2 - MATERIAL BEHAVIOR II

Chair: A. Chutjian, Jet Propulsion Laboratory

3:45 p.m. "New Experiments on the Langmuir Film Balance"
G. Morrison, National Bureau of Standards

4:15 p.m. "The Supercritical Fluid Chromatograph"
T. J. Bruno, National Bureau of Standards

4:45 p.m. "A High Temperature Thermal Conductivity Apparatus"
H. M. Roder, National Bureau of Standards

5:15 p.m. "Cyclic Hardening and Ratchetting of Type 304 Stainless Steel, Experiment and Modeling"
E. Krempl, Rensselaer Polytechnic Institute

THURSDAY, JUNE 18

TECHNICAL SESSION 3 - PLASMA PROCESSES

Chair: T. Zipperian, Sandia National Laboratory

8:00 a.m. "Laser Diagnostics of PACVD Processes for Depositing Hard Face Coatings"
W. Roman, United Technologies Research Center

8:30 a.m. "Developing Diagnostic Methods for Plasma-Particle Interaction"
C. B. Shaw, Idaho National Engineering Laboratory

9:00 a.m. "High Temperature Particle Processing;" and "High Speed Multicolor Pyrometry"
J. F. Elliott, Massachusetts Institute of Technology

9:30 a.m. "Mathematical Modeling of Transport Phenomena in Plasma Systems"
J. Szekely, Massachusetts Institute of Technology

10:00 a.m. BREAK
TECHNICAL SESSION 4 - FLUID MECHANICS/ROBOTICS

Chair: H. Aref, University of California

10:15 a.m.  
"Visualization Techniques for Understanding Particle Dispersion in Turbulent Shear Flows"  
T. R. Troutt, Washington State University

10:45 a.m.  
"Thinning and Rupture of a Thin Liquid Film on a Horizontal Heated Solid Surface"  
S. G. Bankoff, Northwestern University

11:15 a.m.  
"Experimental and Theoretical Studies of Binary Vapor Condensation"  
G. Wilemski, Physical Sciences, Inc.

11:45 a.m.  
Experimental Research in Autonomous Robotics"  
W. R. Hamel, Oak Ridge National Laboratory

12:15 p.m.  
LUNCH

TECHNICAL SESSION 5 - NON DESTRUCTIVE EVALUATION

Chair: W. Roman, United Technologies Research Center

1:15 p.m.  
"An Automated Multiviewing Ultrasonic Technique for Flaw Reconstruction"  
D. K. Hsu, Ames Laboratory

1:45 p.m.  
"Photothermal and Photoacoustic Measurements"  
Jeffrey Fanton, Stanford University

2:15 p.m.  
"An Acoustic Emission Digital Data Acquisition Workstation"  
J. A. Johnson, Idaho National Engineering Laboratory

2:45 p.m.  
"Masking Effects of Microcracks, Voids and Near-Tip Mechanical Behavior on Crack Detection and Characterization"  
J. O. Achenbach, Northwestern University

3:15 p.m.  
BREAK
TECHNICAL SESSION 6 - WELDING PROCESSES

Chair: J. Elliott, Massachusetts Institute of Technology

3:30 p.m. "Instrumentation, Measurement, and Analysis for Controlling Residual Stresses and Metal Movement in Weldments" In-Hwa Chang, Massachusetts Institute of Technology

4:00 p.m. "Sensing and Modeling of Gas Metal Arc Welding for Application to Multivariable Control" H. B. Smartt, Idaho National Engineering Laboratory

4:30 p.m. "Metal Transfer in Gas Metal Arc Welding" Y. S. Kim, Massachusetts Institute of Technology

5:00 p.m. "Multivariable Control of the Gas-Metal Arc Welding Process" Hamid Saedi, Massachusetts Institute of Technology

6:00 p.m. RECEPTION

7:00 p.m. SYMPOSIUM DINNER

FRIDAY, JUNE 19

TECHNICAL SESSION 7 - NON LINEAR PROCESSES

Chair: G. Wilemski, Physical Sciences, Inc.

8:00 a.m. "New Directions in Nonimaging Optics" R. Winston, The University of Chicago

8:30 a.m. "WKB Evolution of Wave Packets" R. G. Littlejohn, Lawrence Berkeley Laboratory

9:00 a.m. "Chaotic Advection: Efficient Mixing of Viscous Fluids" H. Aref, University of California

9:30 a.m. "Structure Formation and Coherence Produced by Mixing in Two-Dimensional Chaotic Fluid Flows" J. M. Ottino, University of Massachusetts

10:00 a.m. BREAK
TECHNICAL SESSION 8 - MATERIAL BEHAVIOR III

Chair: G. Morrison, National Bureau of Standards

10:15 a.m.       "Comminution Within Particle Beds"
                 C. Peterson, Massachusetts Institute of Technology

10:45 a.m.       "Applications of Acoustic Emission Techniques to
                 Cryogenic Experiments"
                 Y. Iwasa, Massachusetts Institute of Technology

11:15 a.m.       "Modeling and Analysis of Surface Cracks"
                 D. M. Parks, Massachusetts Institute of Technology

11:45 a.m.       "Effective Elastic Properties of Cracked Solids"
                 M. Kachanov, Tufts University

12:15 p.m.       CLOSING REMARKS
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH SPEED TOMOGRAPHIC OPTICAL DATA ACQUISITION SYSTEM FOR COMBUSTION RESEARCH</td>
<td>R. Snyder and L. Hesselink (Stanford Univ.)</td>
<td>1</td>
</tr>
<tr>
<td>ELECTRON ATTACHMENT PHENOMENA IN HEATED PLASMAS</td>
<td>A. Chutiian, S. H. Alajajian, and O. J. Orient (California Inst. of Technology)</td>
<td>13</td>
</tr>
<tr>
<td>GaAs/A10.3Ga0.7As, HETEROJUNCTION BIPOLAR TRANSISTORS FOR HIGH-TEMPERATURE (T &gt; 300°C), POWER ELECTRONIC APPLICATIONS</td>
<td>T. E. Zipperian, T. M. Brennan, and B. E. Hammons (Sandia National Lab.)</td>
<td>19</td>
</tr>
<tr>
<td>SEMI-AUTOMATED BURNETT PVT APPARATUS; PROPERTIES OF A GEOTHERMAL WORKING FLUID</td>
<td>D. Linsky, J. M. H. Levelt Sengers, and H. A. Davis (National Bureau of Standards, Gaithersburg)</td>
<td>26</td>
</tr>
<tr>
<td>PROPERTIES OF ELASTOMERIC ALMINATES</td>
<td>S. K. Clark and R. N. Dodge (Univ. of Michigan)</td>
<td>37</td>
</tr>
<tr>
<td>NEW EXPERIMENTS ON THE LANGMUIR FILM BALANCE</td>
<td>G. Morrison (National Bureau of Standards, Gaithersburg), and I. L. Pegg (Catholic Univ. of America/National Bureau of Standards, Gaithersburg)</td>
<td>45</td>
</tr>
<tr>
<td>PHYSICOCHEMICAL APPLICATIONS OF SUPERCRITICAL FLUID CHROMATOGRAPHY</td>
<td>T. J. Bruno (National Bureau of Standards, Boulder)</td>
<td>52</td>
</tr>
<tr>
<td>A HIGH TEMPERATURE THERMAL CONDUCTIVITY APPARATUS FOR FLUIDS</td>
<td>H. M. Roder (National Bureau of Standards, Boulder)</td>
<td>61</td>
</tr>
<tr>
<td>RATCHETTING BEHAVIOR OF TYPE 304 STAINLESS STEEL AT ROOM AND ELEVATED TEMPERATURE</td>
<td>M. Ruggles and E. Krempl (Rensselaer Polytechnic Inst.)</td>
<td>69</td>
</tr>
<tr>
<td>LASER DIAGNOSTICS OF PACVD PROCESSES FOR DEPOSITING HARD FACE COATINGS</td>
<td>W. C. Roman, J. H. Stufflebeam, A. C. Eckbreth, and F. A. Otter (United Technologies Research Center)</td>
<td>77</td>
</tr>
<tr>
<td>DEVELOPING DIAGNOSTIC METHODS FOR PLASMA-PARTICLE INTERACTION</td>
<td>C. B. Shaw, Jr. (EG&amp;G Idaho, Inc.)</td>
<td>85</td>
</tr>
<tr>
<td>HIGH TEMPERATURE PARTICLE PROCESSING</td>
<td>J. F. Elliott, R. E. Spjut, and P. Bolsaitis (Massachusetts Inst. of Technology)</td>
<td>93</td>
</tr>
<tr>
<td>VISUALIZATION TECHNIQUES FOR UNDERSTANDING PARTICLE DISPERSION IN TURBULENT SHEAR FLOWS</td>
<td>T. R. Troutt, R. Hanle, N. Kamalu, F. Wen, and J. N. Chung (Washington State Univ.)</td>
<td>100</td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>STABILITY OF THIN FILMS</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>S. G. Bankoff and S. H. Davis (Northwestern Univ.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL AND THEORETICAL STUDIES OF BINARY VAPOR CONDENSATION</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>M. B. Frish, R. Waterhouse, and G. Wilemski (Physical Sciences Inc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTS IN AUTONOMOUS ROBOTICS</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>W. R. Hamel (Oak Ridge National Lab.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN AUTOMATED MULTIVIEWING ULTRASONIC TECHNIQUE FOR FLAW RECONSTRUCTION</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>D. O. Thompson, D. K. Hsu, and S. J. Wormley (Ames Lab.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIGH-SENSITIVITY OPTICAL PROBE FOR PHOTOTHERMAL MEASUREMENTS</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td>J. T. Fanton and G. S. Kino (Stanford Univ.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN ACOUSTIC EMISSION DIGITAL DATA ACQUISITION WORKSTATION</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>J. A. Johnson, B. A. Barna, and R. T. Allemeier (EG&amp;G Idaho, Inc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRACK CHARACTERIZATION BY SCATTERING METHODS</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>J. D. Achenbach (Northwestern Univ.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSTRUMENTATION, MEASUREMENT, AND ANALYSES FOR CONTROLLING RESIDUAL</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>STRESSES AND METAL MOVEMENT IN WELDMENTS - Especially Reducing Joint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mismatch and Forces Acting on Tack Welds During Butt Welding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-H. Chang, H. Miyachi, and K. Masubuchi (Massachusetts Inst. of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENSING AND MODELLING OF GAS METAL ARC WELDING</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>H. B. Smartt, J. A. Johnson, R. T. Allemeier, N. M. Carlson,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. J. Einerson, and A. D. Watkins (EG&amp;G Idaho, Inc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISIBLE LIGHT EMISSIONS DURING GAS TUNGSTEN ARC WELDING AND ITS</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td>APPLICATION TO WELD IMAGE IMPROVEMENT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. W. Kim, C. Allemand, and T. W. Eagar (Massachusetts Inst. of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NON-LINEARITIES IN WELDING SYSTEM DYNAMICS</td>
<td>184</td>
<td></td>
</tr>
<tr>
<td>D. E. Hardt, C. Doumanidis, and M. Hale (Massachusetts Inst. of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEW DIRECTIONS IN NONIMAGING OPTICS</td>
<td>198</td>
<td></td>
</tr>
<tr>
<td>R. Winston, X. Ning, J. J. O'Gallagher, and W. T. Welford (Univ. of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WKB EVOLUTION OF WAVE PACKETS</td>
<td>206</td>
<td></td>
</tr>
<tr>
<td>R. B. Littlejohn (Lawrence Berkeley Lab.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAOTIC ADVECTION: THE EFFICIENT STIRRING OF VISCOUS LIQUIDS</td>
<td>209</td>
<td></td>
</tr>
<tr>
<td>H. Aref and S. W. Jones (Univ. of California, San Diego)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIXING: STRETCHING, BREAKUP, AND CHAOS</td>
<td>218</td>
<td></td>
</tr>
<tr>
<td>J. M. Ottino (Univ. of Massachusetts, Amherst)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
COMMINUTION WITHIN PARTICLE BEDS
C. R. Peterson (Massachusetts Inst. of Technology) ................. 226

APPLICATION OF ACOUSTIC EMISSION TECHNIQUES TO CRYOGENIC EXPERIMENTS
Y. Iwasa (Massachusetts Inst. Technology) ......................... 236

MODELING AND ANALYSIS OF SURFACE CRACKS
D. M. Parks (Massachusetts Inst. of Technology) ................. 244

SOLIDS WITH MULTIPLE CRACKS AND SOME RELATED PROBLEMS
M. Kachanov (Tufts Univ.) ........................................... 252

* * * * *

LIST OF PARTICIPANTS .................................................. 259
HIGH SPEED TOMOGRAPHIC OPTICAL DATA ACQUISITION SYSTEM FOR COMBUSTION RESEARCH

Ray Snyder
and
Lambertus Kesselink
Department of Aeronautics/Astronautics
and Electrical Engineering
Stanford University
Stanford, Ca 94305-2186

June 6, 1987

1 Introduction

Optical tomography provides a means for obtaining spatially resolved measurements from line-of-sight data. Classical flow visualization approaches such as shadowgraphy, schlieren and interferometry provide information about physical observables averaged along the path of the probe beam. Locally resolved measurements cannot be obtained from such data sets, except for axially symmetric or two-dimensional flows. On the other hand, scattering techniques provide space resolved imagery by collecting probe radiation outside the illumination plane. Unfortunately, scattering cross sections are usually small implying weak signals and large laser systems are often required to make measurements over an extended area. In hot exhaust plumes or combusting flows collecting optics may not always be placed outside the illumination plane and line-of-sight techniques may be preferred.

Tomography has been applied to medical and industrial applications [1]. Particularly in medicine, tomographic diagnostics are well developed. Since medical objects tend to be stationary, data acquisition times need not be very rapid and sequential acquisition of projections is sufficient. In fluid mechanics no such luxury is usually allowed, since the most interesting flows
are unsteady and rapidly developing. This requires that tomographic data acquisition systems have a very short response time and all projections must be collected in such a short time that the flow is essentially frozen.

In the past, optical acquisition implementations have had a long time response, allowing only time averaged or stationary flows to be examined. These approaches, unfortunately, are not suitable for investigation of turbulent combusting flows.

In this paper we discuss a new optical architecture capable of providing 36 equally spaced projections about a 180 degree arc in 300 μsec. Each projection passes through a three inch diameter, three inch high cylinder, allowing a full three-dimensional reconstruction of the flow in that region. Holographic interferometry allows direct measurement of optical phase along the pathlength. From these data, the flow may be reconstructed using any of the available reconstruction techniques.

In 1917, Radon showed that perfect image reconstruction can be achieved using an infinite number of perfect projections. Noisy data and a limited set of projections make unambiguous reconstruction impossible, and artifacts are introduced in the image. The specifics of the artifacts depend on the reconstruction technique used, the noise level and the number and spacing of the projections. The most popular reconstruction technique is the convolution back projection; it is an efficient and fast method. Reconstruction quality is good and artifacts may be reduced by post processing.

In this paper we first describe the design of the optical apparatus, followed by an outline of the experiment, the data reduction scheme and the results obtained with this method.

2 Modeling Studies

The quality of the final data is dependent on many factors. Among these are the signal to noise ratio, the resolution of the optical components, the sensitivity of the software to measurement errors, the complexity of the flow being measured and the quantity of the measurements.

The spatial resolution attainable is most limited by the spacing between views in recording. From Fourier analysis of the general tomographic procedure, ignoring all other aspects of the problem, the resolution is inversely proportional to the number of views; the more views, the better the resolution. Of course, any real experiment must be limited in the number of views by external constraints. We simulate the measuring process using numerical
Figure 1: Reconstructions of coflowing jet using 20, 30 and 40 projections

projections through a 2 dimensional data field obtained by photographing the scattering from a thin sheet of light illuminating a smoke seeded flow. The Reynolds number of the flow is approximately 3000. From a set of 20, 30 and 40 equally spaced projections to which Gaussian noise is added, reconstructions are obtained using a common algorithm. The reconstructions are shown in Figure 1 with the original image for comparison. Note that with 20 projections (left), some significant detail is ambiguously resolved, whereas with 40 projections (right), all the significant features of the flow in the original data are clearly reconstructed. Quantitative study of the reconstructions shows continuous improvement in the mean square error criterion as the number of views is increased, comparing the reconstruction to the original, but with small gains with increased views after 30 views are used. Close qualitative inspection shows that much of the error is in ringing artifacts associated with the reconstruction algorithm and that with 30 views all the important edge detail is resolved. With fewer views the edge detail may be resolved, but in an ambiguous fashion, so that there may be spurious features.

With a less complicated flow, in particular one with a lower Reynolds number, good qualitative and quantitative reconstructions can be obtained with fewer views; we obtain good resolution in a simulation of a 1000 $Re$ flow with 20 views.

The ringing artifacts can be suppressed by filtering the image. This causes a significantly lower mean square error for reconstructions with any number of views. Repeated filtering will continually reduce the error figure,
but often will not improve the image "quality". In fact, the perceived quality may decline.

3 Outline of the Experiment

The flow field under consideration is a coflowing jet with and without combustion. Line-of-sight holographic interferometry measurements are made of species concentration in the absence of combustion, and temperature with combustion. In particular, it is the objective of the study to investigate the dependence of the interface surface between the two fluids on the Reynolds number and to compare these results with measurements made with smoke seeded flows.

Line-of-sight interferometric measurements are made of index of refraction when viewing the object from many different directions. The index of refraction is directly related to the gas density through the well known Gladstone-Dale relationship; in combusting flows the index is a function of temperature as well as concentration, and for small index variations between the mixing gases, temperature can be determined instead of density.

The optical configuration of the system is shown in Figure 2. An Ar+ laser beam is passed through beam shaping optics and is incident on a small cylindrical mirror. The mirror reflects the light over an angle of 220 degrees and towards 36 focusing mirrors. Each mirror collimates and directs the beam towards the center of the object from 36 different directions uniformly divided along an arc of 180 degrees. The light transmitted through the object is incident on another array of 36 mirrors, oriented in a similar configuration about the object. The latter mirrors focus and direct the beams towards a cylindrical recording plane. A second laser beam and optical arrangement similar to the one described for the object wave is used to illuminate the film plane with a reference beam. A holographic interferogram is recorded by superimposing two holograms, one without flow and one with flow. The developed holographic interferogram is illuminated with the reference wave in a similar setup as was used during recording to produce a fringe pattern which is recorded on film for subsequent data processing. After development the fringe pattern on the film is digitized with a PDS microdensitometer.

Based upon the simulations, the experimental apparatus has been designed to allow 36 views through a sample volume, 7.5 cm in diameter, 7.5 cm high with .5 mm resolution; all views being obtained in 300 μsec. With a 1" (25mm) diameter inner jet, we measure coflowing jet flows with
Figure 2: Optical Configuration
4 Data Reduction

Referring to Figure 2, each projection contains information about the index-of-refraction integrated along the path (the optical path length). The interferogram encodes intensity as a function of phase which is recorded holographically. To recover phase from the interference pattern a tilt in the reference beam is introduced between exposures. This assures phase encoding as a deviation of fringes from a straight line in a finite fringe interferogram. Such interferograms may be analyzed using Fourier approaches [2]. Fourier techniques are preferred over fringe tracking, because all intensity information in the patterns is used, and not just intensity extrema.

Processing proceeds by digitizing the fringe pattern as shown in Figure 3) using a microdensitometer on a 512x512 grid of points or pixels. Amplitude resolution is ten bits. Subsequently, the 1-D Fourier transform is taken in a direction perpendicular to the fringe spacing in the background pattern. To reconstruct a two-dimensional cross section a single line in the interferogram is processed; three-dimensional reconstructions require analysis of two-dimensional projections. In Fourier space the optical path length infor-
mation is in contained in spatial frequency components grouped around the
frequency of the background bias (Figure 4). This portion of the spectrum is
separated from the rest of the spectrum by filtering with a Hanning window,
and is shifted to the DC position (Figure 5). The inverse Fourier transform
then directly yields the phase difference between the two passes through the
test section (Figure 6). An automated procedure is used to unwrap $2\pi$ phase
jumps. Using, in this case 18 projections, the flow is reconstructed. Arti-
facts are minimal, but image quality may be improved by median filtering
which removes isolated noise spikes.

5 Results

The result of the measurements is shown in Figure 7. The index of refraction
is plotted throughout the flow cross section as a perspective view. The
signal to noise ratio is excellent, which may be seen in Figure 8. Here
concentration of Helium is plotted as a function of position along a line in
the flow. The background value of 0.0 corresponds to air and 1.0 to pure
helium. It is noteworthy here that the central peak is not scaled, but directly
computed. This shows that the numerical representation is very accurate
indeed, the error being less than 3%. Another representation of the flow is
Figure 5: Filtered Fourier Pattern

Figure 6: Zero Fringe Pattern
shown in Figure 9, where each contour line corresponds to an increase of .15 in normalized density; zero corresponds to air and one to helium. This result, displayed as a grayscale image as shown in Figure 10, may be directly compared with a cross sectional image shown in Figure 11 obtained by Mie scattering in a similar experiment. The inner fluid is seeded with smoke and a sheet of laser light illuminates a cross section perpendicular to the mean flow direction. It is interesting to note the correspondence between the geometrical shapes of these two results. However, diffusion of helium into air proceeds at a much faster rate than particulate diffusion, which explains why the interface between the two fluids is much more diffuse in the tomographic image than in the smoke visualization.

6 Discussion

The results presented here are the first instantaneous tomographic representations of a non-stationary flow. The time response is very fast, approximately 300μsec exposure time with a 1 Watt argon laser, allowing turbulent flows to be measured. Spatial resolution is very good using 36 projections and the signal-to-noise ratio is excellent with errors less than 3%. Spatial resolution in the flow direction is limited to 1.5 mm for the present con-
Concentration of Helium

Figure 8: Concentration of Helium in Coflowing Jet

Figure 9: Contour Plot of Helium Concentration
Figure 10: Gray Scale Image of Reconstructed Flow Section

Figure 11: Mie Scattering Image of Reconstructed Flow Section
Figuration using a jet velocity of 6 m/sec. Exposure time may be further reduced by using more sensitive film. Spatial resolution in the transverse direction is less than 1.5 mm.

Currently tomographic imaging of the transition region in the jet is being pursued. Here, a full three-dimensional representation of the instantaneous density field is to be obtained. Extension to combusting flows is immediate, however we need to assume a priori knowledge about the temperature field in order to relate optical phase directly to concentration. Or, alternatively, temperature may be measured in a flow where both components have the same index of refraction. With a slight modification, absorption tomography may be directly carried out by using a tunable dye laser. By probing the flow near a resonance line of either one of the fluids or a combusting product, species concentration may be measured.

7 Acknowledgement

This research was supported by DOE under grant DE-FG03-85ER13312.

References


ELECTRON ATTACHMENT PHENOMENA IN HEATED PLASMAS

A. Chutjian, S. H. Alajajian and O. J. Orient
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109 USA

ABSTRACT

Recent JPL experimental and theoretical results of electron attachment to molecules at ultralow energies are discussed. Specifically, electron attachment to $F_2$, and temperature dependencies of attachment rate constants in $CFCI_3$ and $CCl_4$ are used to illustrate some interesting features. Extensions to several important areas involving plasma-surface interactions are outlined.

INTRODUCTION

Gases charged by cathode discharges that are driven thermally, by electrons from a hot filament, or by microwave and rf discharges are being used in a wide and expanding variety of industrial applications. Two prominent applications are in combustion[1] and semiconductor processing[2,3]. The combustion applications deal with reduction or enhancement of soot formation and electron densities (the two may be coupled) to reduce radar signatures in a rocket exhaust, for example. Plasma-assisted etching is used by most, if not all, semiconductor manufacturers. Typical additives to these plasmas are halogenated carbon compounds ($CF_4$, $CCl_4$, $C_2F_2Cl_2$, $C_2F_3Cl$), $SF_6$, and dilute mixtures with $F_2$ and $Cl_2$.

A dearth of knowledge exists on cross sections and rate constants, and their temperature dependencies, of important phenomena occurring in these chemically active plasmas. Thermal electron attachment to the additive molecule plays a key role. Cross sections for attachment are not known in nearly all cases, since the mean electron energies are very low (less than 0.5 eV, or 3000 K), and measurements are difficult. Furthermore, it is only recently being noted[4] that cross sections or rate constants measured at room temperature under single-collision or multiple-collision (swarm) conditions may not apply to plasmas operating at 500–1000 K where vibrational excitation of the target will occur, or where the electron energy distribution function may deviate from Maxwellian, or where equilibrium conditions may not prevail.

In the following we discuss several recent results of our work in electron attachment at ultralow electron energies. We also indicate extensions of this and related work at JPL to areas of plasma-surface interactions using well-characterized beams of fast, high-flux oxygen and hydrogen atoms. These interactions include the phenomena of chemical vapor depositions (CVD) using $H$-atom irradiation of a substrate to form $sp^3$ diamond rather than graphite; the formation of low-$T$ oxides on $Si$ and $GaAs$ surfaces using $O$-atoms; and oxygen-atom injection to form the $La_{2-x}Ba_xCuO_{4-y}$ and $Y_{0.87}Ba_{1.55}Cu_{2.5}O_{y}$-type, high-$T_c$ superconductors. Extensions to the use of laser opto-galvanic spectroscopy in the study of electron-assisted CVD plasmas is also discussed.

1 Electron Attachment Phenomena at Ultralow Electron Energies

Using the krypton photoionization method developed by us at JPL[5], we are able to measure attachment cross sections over a range of electron energies and resolutions not possible by any other technique. We have, by this method, shown that the attachment cross section $\sigma_A(\epsilon)$ at energies $\epsilon$ below 10 millielectron volts (meV) is dominated by the s-wave threshold law, in which $\sigma_A(\epsilon) \sim \epsilon^{-1/2}$; that is, it becomes infinite just at zero electron energy. In addition to publishing results on about twelve molecules over the past year[6–9], we have completed a difficult series of measurements on $F_2[10]$, and have completed calculations of cross sections and rate constants in $CFCI_3$ and $CCl_4[11]$, analogous to our calculations in $SF_6[12]$. We present below, by way of highlight, these last two results.
1.1 Electron Attachment in $F_2$

The halogen molecules are frequently used in plasma etching, and as the working substance in a range of lasers ($KrF$, $F_2$, and $I_2$ lasers). From an applied engineering point of view, cross section measurements in $F_2$, $Cl_2$ and $I_2$ provide part of the data base needed to model the plasmas, to effect etching control, or to optimize laser efficiency. Molecular fluorine was interesting from the basic physics point of view as well. Several very accurate theories\cite{13,14} have predicted that the attachment cross section in $F_2$($e + F_2 \rightarrow F + F^-$) should be $p$-wave in nature, or $\sigma_A(e) \sim e^{1/2}$; i.e., cross section vanishing at zero energy. Earlier measurements\cite{15,16} taken at very low energy resolution, and with considerable uncertainty in the energy scale, appeared to show an opposite trend. We undertook measurements on this corrosive molecule, and our results are shown in Fig. 1. Our measured cross sections (solid line) are obtained at a resolution of 6

\[ F^- / F_2 \]

\[ 10^{-14} \]

\[ 10^{-15} \]

\[ 6 \text{ meV} \]

\[ \text{ATTACHMENT CROSS SECTION (cm}^2\text{)} \]

\[ \text{ELECTRON ENERGY (meV)} \]

Figure 1. Electron Attachment Cross Sections in $F_2$. Present experimental results are shown as solid line. Theoretical curves are from Refs. 13 (----) and 14 (--.--.--).

meV (FWHM) and are uncertain to 25%, whereas the swarm data\cite{15} (filled circles, 40 meV resolution) and beam data\cite{16} (open circles, 100 meV resolution) are unreliable below about 40 meV energy. Present results clearly show an attachment cross section rising as $e^{-1/2}$, or as an $s$-wave cross section. Thus the discrepancy is between a quantity vanishing at zero energy (theory), or diverging (experiment)! Two possible causes of this difference are the neglect, by theory, of small nonadiabatic couplings between a higher-energy $^2\Sigma^+$ electronic state, or Coriolis coupling between a nearby $^3\Pi_g$ state and the $\Pi_g$ rotations, to give a $^1\Sigma_g^+$ component to the $F_2^- \, ^2\Sigma_g^+$ ground state. This behavior of cross section has a marked effect on the behavior of the attachment rate constant $k(\tilde{e})$ with mean electron energy $\tilde{e}$ or temperature $T$.

We would like to study the $Cl_2$ molecule as well, for which experiment now shows a decreasing cross section towards zero energy\cite{15}. No theory is available.
1.2 Calculations of $k(\epsilon)$ vs Temperature

We undertook a theoretical study of attachment cross sections and rate constants $k(\epsilon)$ in the molecules $\text{CFCI}_3$ and $\text{CCl}_4$, for which we had measurements. Our curiosity was aroused by measurements, using the flowing afterglow technique, by Smith et al. of $k(\epsilon)$ vs $T$ which showed very different behaviors for these two molecules. Their experimental results are shown in Figs. 2 and 3 (filled circles), and present calculations by the solid line. Theoretically, these two molecules behave quite similarly, and we were unable to account for the rise in $k(\epsilon)$ with $T$ for $\text{CFCI}_3$, as was true also in $\text{SF}_6$. We sent our results to Prof. R. W. Crompton of the Australian National Univ. and asked him to measure one high-temperature value of $k(\epsilon)$ in $\text{CFCI}_3$ for us. His result is given by the filled square at 500 K in Fig. 2, in very good agreement with our calculation. At this point, we can only surmise that the discrepancy lies in the fact that there are nonequilibrium or non-Maxwellian conditions prevailing in the flowing afterglow technique. If so, this is an important result, given that equilibrium is the underlying assumption of the flowing afterglow technique, and much data have been obtained using it. Some independent confirmation of this may be forthcoming from the work of Ratliff et al. in which they find that, in a flowing helium afterglow, the electron flow velocity exceeds the helium flow velocity by a factor of five. From an applications standpoint, the results of Figs.

![Figure 2. Attachment Rate Constants in $\text{CFCI}_3$.](image2)

![Figure 3. Attachment Rate Constants in $\text{CCl}_4$.](image3)

Present calculations are shown as solid line.
2 and 3 would have bearing on whether these or other molecules would be effective, at higher \( T \), in soot formation, or in radar concealment (reduction of free electron density) of rocket exhausts.

We plan to publish our calculations in these molecules jointly with the Australian National Univ. group when their measurements on \( \text{CFCI}_3 \) and \( \text{CCl}_4 \) are completed. We would also plan on continuing the calculations in other molecules (\( \text{HI}, \text{DI} \) and \( \text{CH}_2\text{Br}_2 \)) to verify trends there.

2 Plasma-Surface Interactions

There have been several exciting developments in the area of plasma-surface interactions. Three which we wish to emphasize are (1) use of chemical vapor depositions (CVD) to grow diamonds on substrates[21], (2) possibility for growth of uniform, low-\( T \) oxides[22] on \( \text{Si} \) and \( \text{GaAs} \) semiconductors using beams of oxygen atoms[23], and (3) the growth of new, high-temperature superconductor films of the type[24] \( \text{Y}_{0.87}\text{Ba}_{1.53}\text{Cu}_3\text{O}_y \).

In each area, atomic interactions with a surface are believed to play a decisive role. In (1), \( \text{H} \)-atoms binding to \( sp^3 \)-hybridized carbon atoms are believed to maintain this hybridization so that diamond-like carbon, rather than graphite, can be deposited. In (2), \( \text{O} \)-atoms impinging on a \( \text{Si} \) or \( \text{GaAs} \) surface lead to oxidation. It is desirable to have low surface temperatures in order to reduce the formation of defects, and minimize diffusion into the substrate. In (3) control of the oxygen content in the superconductors is believed to have a strong effect on the temperature and width of the normal-to-superconducting transition region.

Over the past several years we have been developing at JPL a high-flux, fast-atom beam source for Shuttle glow studies[23]. The source is currently in need of funding for a laser to effect detachment of the fast negative ion. It is shown schematically in Fig. 4 for \( \text{O} \)-atom generation, although \( \text{H} \)-atoms can be generated.

![JPL VARIABLE ENERGY, HIGH FLUX ATOMIC OXYGEN SOURCE](image)

Figure 4. Schematic of the JPL Fast-Atom Source suitable for generating clean atomic beams of \( \text{O}(^3\text{P}) \) or \( \text{H}(^3\text{S}) \).
as well by changing the target feed gas. Details of the source will be given at the Symposium. It is capable of generating fast (3-20 eV), high flux \((10^{14} \text{ atoms/cm}^2\text{s})\) ground state \([O(3P) \text{ or } H(2S)]\) atoms. As such, one now may (1') study the effects on diamond growth of a beam of \(H\)-atoms on a methane-exposed substrate, (2') using \(O\)-atoms, form low-temperature \(SiO\) and \(GaAs-O\) semiconductors which can have potentially lower defect and interface states, (3') using \(O\)-atoms, control the oxide content of the thin film superconductors, and study the effect of transition-region temperature and width with \(O\)-atom exposure.

Furthermore, in connection with (1) above, we have established a strong desire between JPL and GTE Labs (Waltham, MA)[25] to collaborate on the study of the basic plasma properties of the CVD phenomenon using laser opto-galvanic spectroscopy. A combination of the experimental capabilities at JPL, and the strong modeling/swarm-theory presence at GTE Labs would help shed light on the types of species present, and their role, in the CVD and electron-assisted CVD plasma used to generate surface coatings.

The capability also exists, with minor changes in our photoionization apparatus, to investigate attachment properties of molecular clusters formed in high-pressure discharges, such as the CVD plasma. As a test case, we would investigate the \(SF_6-Kr_n(n=1-4)\) cluster, analogous to the \(SF_6-Ar_n\) cluster[26].

This work was sponsored by the Department of Energy, and was carried out at the Jet Propulsion Laboratory, California Institute of Technology, through agreement with NASA.

References


GaAs/Al$_{0.3}$Ga$_{0.7}$As, HETEROJUNCTION BIPOLAR TRANSISTORS
FOR HIGH-TEMPERATURE (T > 300 °C), POWER ELECTRONIC APPLICATIONS

T.E. Zipperian, T.M. Brennan, and B.E. Hammons
Division 1141
Sandia National Laboratories
Albuquerque, NM 87185

ABSTRACT

Solid-state devices formed from compound semiconductor materials like GaAs, GaP, SiC, and (Al,Ga)As have long been viewed as candidates for use in electronic circuits functioning at temperatures greater than 300 °C. To address the specific needs of power semiconductor devices operating simultaneously at high currents, voltages, and high temperatures, heterojunction devices formed from combinations of GaAs and (Al,Ga)As have recently been proposed. These novel heterojunction structures display reduced resistive and voltage parasitics when compared to wide-bandgap GaP or SiC homojunction diodes without seriously compromising control of thermally generated leakage currents. In this study a prototype, low-power, (Al,Ga)As/GaAs, heterojunction bipolar transistor (HBT) is described which has demonstrated excellent electrical characteristics in the 300 to 400 °C temperature range. At 350°C, the HBT has a common-emitter current gain of 14 (V$^c_e$ = 5V, I$^e$ = 10 mA) and collector-base leakage of 6.4x10$^{-2}$ A/cm$^2$ (V$^c_b$ = 5 V). These studies and others imply that a heterojunction, (Al,Ga)As/GaAs, power semiconductor electronics technology is feasible in the near future.

INTRODUCTION

A variety of applications in jet engine control instrumentation, well logging, electronics for space probes, and recently, power conversion, conditioning, and control electronics for space-based power systems, are driving the development of solid-state components capable of operating for sustained periods of time at high temperatures (T > 300 °C). Devices formed from wide-bandgap semiconductor materials like GaAs, GaP, and SiC, have long been viewed as potential candidates for use in these difficult environments. Temperature records for solid-state devices include the successful operation of GaP/(Al,Ga)P, heterojunction bipolar transistors (HBTs) [1] and heterojunction semiconductor controlled rectifiers (HSCRs) [2] for limited periods of time at 550 °C, and the demonstration of useful rectification in SiC, p-n diodes [3,4] at 500 °C for 1000 h, with no apparent degradation. Similar SiC, p-n diodes [3] have been used as particle detectors to temperatures as high as 700 °C.

A particularly severe high-temperature applications area, and potentially one where use of alternative materials other than Si may have its greatest impact, is in power conversion and conditioning. Almost all power devices, whether bipolar or field-effect, rely on p-n junctions for either operation or protection [5]. A major problem in solid-state devices is the fact that reverse leakage currents in p-n junc-
tions increase exponentially with temperature, leading eventually to component failure. This problem is particularly severe in power devices where junction leakage causes significant internal power dissipation when a structure is biased in a blocking state. Many switching devices also have internal gain mechanisms [5] which amplify junction leakage, further compounding the problem. Reverse leakage exponentially decreases as the energy bandgap of the material used to form the junction increases. Use of wide bandgap semiconductors instead of Si directly addresses this and other concerns. A recent survey [6] of the use of compound semiconductors in power semiconductor electronics describes this in greater detail.

Although wide-bandgap materials are required for high-temperature operation, they cause a number of other problems for power semiconductor device design. In general, electron and hole mobilities decrease with increasing bandgap (GaAs a notable exception), and metal-semiconductor barrier heights usually increase with bandgap. These tendencies increase resistive parasitics associated with the substrate material and contacts in wide-bandgap semiconductor devices and increase internal power dissipation at high currents. Forward p-n junction voltage drops are also higher in wide bandgap materials. For example, at room temperature comparable Si and GaP homojunction diodes forward biased to identical current densities in the range of 10 A/cm$^2$ will have forward voltage drops of approximately 0.75 and 1.8 V, respectively. Although these values decrease with temperature, internal power dissipation increases with increases in the forward voltage drop.

A solution was proposed [7] which eases the problems associated with use of wide-bandgap materials in high-temperature power junction devices. Lattice-matched heterojunctions can be formed between wide-bandgap and narrow-bandgap materials in systems like (Al,Ga)As/GaAs, permitting one to use wide-bandgap materials selectively as needed to control thermally generated reverse leakage currents. With appropriate device design, narrower-bandgap material can be used for the substrate and in all contacting areas to decrease bulk and contact [8] resistive parasitics. Band-offset effects [9] also cause the forward voltage drops of narrow-bandgap/wide-bandgap, p-n heterojunctions to be comparable to homojunctions formed solely in the narrow-bandgap material.

Initial studies of (Al,Ga)As/GaAs, heterojunction diodes [7] have shown that rectifiers with reverse leakage current densities less than $5 \times 10^{-2}$ A/cm$^2$ (VR - 10 V) at 400°C can be demonstrated. These same structures have room temperature forward voltage drops of 1.25 V at forward current densities of 10 A/cm$^2$, a value comparable to p-n diodes in bulk GaAs. All ohmic contacts in these devices were to the narrower-bandgap GaAs material and the structures were grown on n$^+$ GaAs substrates. This significantly reduces resistive parasitics when compared to GaP structures. This reverse leakage is approximately three orders of magnitude lower than for a comparable Si diode, two orders of magnitude lower than for a GaAs homojunction structure, and one order of magnitude higher than for a GaP device.

These data show that for a slight reduction in the high-temperature performance ultimately available from devices formed in GaP or SiC, (Al,Ga)As/GaAs, heterojunction diodes offer significant improvements in both resistive and voltage parasitics. Weighing considerations of material availability, expected and demonstrated device performance, and the available technology base, heterojunction structures in the (Al,Ga)As/GaAs materials system are the best choice for the near term (five year) commercialization of junction devices for high-temperature (T > 300°C) power semiconductor applications.

A first-order extension of the heterojunction concepts utilized in the diodes of reference [7] to a three-terminal, high-temperature switching device, is to incorporate (Al,Ga)As selectively in the collector of a GaAs bipolar transistor. Such a structure is shown in cross section in Fig. 1. Heterojunction emitters and collectors have previously been utilized in (Al,Ga)As/GaAs bipolar structures [10,11] to increase current gain and frequency response and to reduce minority carrier storage effects when the transistor is driven into saturation. (Al,Ga)As is used selectively in the collector of this device solely to control thermally generated leakage currents; no control of minority carrier injection or storage is intended. The AlAs concentration was chosen as 30 % in the (Al,Ga)As layer because prior heterojunction diode studies [12] have shown this composition to be an effective compromise between improved high-temperature electrical characteristics and degraded chemical stability.
of the Al-bearing material. The growth, fabrication, and electrical evaluation at temperatures up to 450 °C of transistors similar to that shown in Fig. 1 are described below.

EXPERIMENTAL PROCEDURE

The n⁺p⁺n⁺, (Al,Ga)As/GaAs HBT of Fig. 1 was formed by computer-controlled molecular beam epitaxy (MBE) in a Varian Gen II system. The epitaxial layers were grown on (100), Sumitomo, Si-doped GaAs substrates. Polished substrates were cleaned [13], mounted with In on molybdenum carriers, and introduced into the machine. Substrate temperatures were monitored by a spring loaded thermocouple in contact with the back side of the sample holder and also by a single wavelength optical pyrometer looking normal to the surface through a quartz window. All GaAs layers were grown at 600 °C and at a growth rate of 1 μm/h. The Al₀.₃Ga₀.₇As layer was grown at a temperature of 700 °C and also at a growth rate of 1 μm/h. The V/III flux ratio was maintained at ~1.75 at all times. Be and Si were used as the p and n-type dopants, respectively. The epitaxial growth of all electrically active layers was initiated on an n⁺ GaAs buffer followed by a GaAs/AlAs short-period superlattice (SPS) which was used as an active junction smoothing layer [14]. When an Al₀.₃Ga₀.₇As/GaAs interface was encountered, the growth was halted for 1 minute to allow the substrate temperature to change.

Once the composition, resistivity, and thickness of all layers were formed by MBE, the emitter and base regions were isolated by mesa formation and electrical contact was made to all active layers. Thermally evaporated Au-Be/Au was used as the p-type metallization and electron-beam evaporated Au-Ge/Ni/Au was used for n-type contacts. Individual devices were mounted in ceramic flat-packs using a silver-loaded polyimide adhesive and contacted using thermocompression-bonded 1.5 mil Au wire. The contacting systems and packaging techniques are inappropriate for extended operation at high temperatures; however, they are sufficient for an initial evaluation of both the device configuration and material limitations.

RESULTS

The n⁺p⁺n⁺, (Al,Ga)As/GaAs HBT shown in Fig. 1 was tested at temperatures from 23 to 450 °C. Common-emitter output characteristics for the device at four temperatures are shown in Fig. 2 and collector-base leakage current as a function of collector-base voltage with the emitter open circuited, at seven temperatures, are shown in Fig. 3. This transistor had an emitter and collector junction area of 3.1x10⁻⁴ and 1.56x10⁻³ cm², respectively. At room temperature the transistor is observed to have a common-emitter, small-signal current gain of 23 (VᵧC = 5 V, IᵧC = 10 mA) and a collector-base leakage of 1.3x10⁻⁷ A/cm² (VᵧC = 5 V). Not shown, the collector-emitter breakdown voltage with the base open-circuited was 40 V. As temperature increases to 300 °C, at the same bias values, the current gain monotonically decreases to a value of 16 and the collector-base leakage increases to 3x10⁻³ A/cm². By 400 °C the collector-base leakage has increased to 4x10⁻¹ A/cm² and is comparable to the base drive; the device is starting to thermally fail.

Common-emitter collector current as a function of base-emitter voltage at four temperatures is shown in Fig. 4. Collector current is observed to increase exponentially with base-emitter voltage up to a value of approximately 1 mA (3.2 A/cm²) at which time the increase with applied bias slows. The flattening of the characteristics is the result of resistive parasitics in the base of the transistor. Structural improvements reducing these parasitics would increase the collector current density at which the flattening occurs.

The I-V data detailed above show that the n⁺p⁺n⁺, (Al,Ga)As/GaAs HBT has usable gain characteristics with low collector-base leakage over the full temperature range from 25 to 400 °C. These data compare favorably with characteristics for both homo-
Emitter Contact

Base Contact

Collector Contact

\( n^+ \) GaAs
\( 2 \times 10^{18} \text{cm}^{-3}, 0.5 \mu \text{m} \)

\( p \) GaAs
\( 1 \times 10^{17} \text{cm}^{-3}, 0.5 \mu \text{m} \)

\( N^- \) Al\(_0.3\)Ga\(_0.7\)As
\( 7 \times 10^{15} \text{cm}^{-3}, 3.0 \mu \text{m} \)

\( n^+ \) GaAs/AlAs SPS
\( 2 \times 10^{18} \text{cm}^{-3} \)
10 periods, 25Å/25Å

\( n^+ \) GaAs Buffer
\( 2 \times 10^{18} \text{cm}^{-3}, 1.0 \mu \text{m} \)

\( n^+ \) GaAs

Substrate

Fig. 1--Structure and equivalent circuit of a prototype, \( n^+pN^- \), molecular beam epitaxial (MBE), GaAs/GaAs/Al\(_0.3\)Ga\(_0.7\)As, heterojunction bipolar transistor (HBT). Capital letters in the device descriptor (\( n^+pN^- \) for example) signify the wider bandgap material in the heterojunction. The emitter and collector areas of this device are 3.1x10^-4 and 1.6x10^-3 cm², respectively.

Fig. 2--Common-emitter output characteristics of the \( n^+pN^- \), GaAs/GaAs/(Al,Ga)As HBT at four temperatures. The base current step for all four traces is 200 \( \mu \text{A/step} \). The collector current density range covered by these characteristics is approximately 0 to 80 A/cm².
Fig. 3--Collector current vs. collector-base voltage with the emitter open-circuited for the n+pN+, GaAs/GaAs/(Al,Ga)As HBT at seven temperatures.

Fig. 4--Common-emitter collector current vs. base-emitter voltage for the n+pN+, GaAs/GaAs/(Al,Ga)As HBT at four temperatures.
junction, GaAs bipolar transistors [15,16] and HBTs composed entirely of (Al,Ga)As [17]. (Al,Ga)P/GaP HBTs [1] have demonstrated higher temperature operation (550 °C) with significantly lower current gains and much higher resistive parasitics.

CONCLUSION

The electrical characteristics described above demonstrate that judicious use of wide bandgap (Al,Ga)As material in critical device regions is sufficient to allow heterojunction bipolar transistors to be fully functional in the 300 to 400 °C temperature range. Junction leakage in these devices is inferior to devices formed in wider bandgap materials like GaP or SiC. However, for power semiconductor devices this penalty is more than offset by the improved characteristics, reduced parasitics, and larger technological base (Al,Ga)As/GaAs offers over other materials.

Significant problems still remain. No compound semiconductor device functional at high temperatures has yet been scaled to high current and voltage operation. A near term goal of our program is the demonstration of 200 V, 20 A, diodes, HBTs and HSCRs, functional at 350 °C. In addition to power scaling considerations, thermally activated chemical processes such as diffusion, electromigration, and corrosion, and the physical, mechanical, and electrical degradation they cause, place an effective maximum temperature limit on the operation of any solid-state device technology once structures electrically capable of high-temperature operation are determined. No failure mechanism [6] identified to date represents a fundamental impediment to reliable, long-lived electronics operating in the 300 to 400 °C range. Even with these challenges, the trend is clear: functional solid-state power electronics operating at temperatures far above the current Si limit is becoming a possibility.

ACKNOWLEDGMENT

This work was sponsored by the U.S. Dept. of Energy under contract DE-AC04-76-DPO0789. The authors wish to thank T.J. Drummond for fruitful discussions during the course of this work and also recognize the expert technical assistance of T.A. Plut, R.B. Caldwell, and J.B. Snelling.

REFERENCES


ABSTRACT

A Burnett-isochoric facility has been constructed in which the equation of state of fluids and fluid mixtures can be obtained in automatic fashion along paths of constant volume. The automated working range is from 273-473 K in temperature, and up to 10 MPa in pressure. We describe how the automation of pressure and temperature control and measurement has been achieved. The apparatus has been used for the determination of the thermodynamic properties of a working fluid mixture for a binary geothermal power cycle. The performance of the apparatus is evaluated by means of the data obtained. Plans for further improvement of range and accuracy are outlined.

INTRODUCTION

Optimization of the performance of thermodynamic working cycles, such as are used in power generation, heating, cooling and waste-heat utilization, requires knowledge of the thermodynamic properties of the working fluid. This information is obtained through measurement of the equation of state of the fluid concerned. In order to obtain the desired thermodynamic derivatives with sufficient accuracy, the basic pressure-volume-temperature (PVT) data must be measured with utmost accuracy. This requires painstaking, tedious effort of long duration, particularly if the working fluid happens to be a mixture. It appears sensible to alleviate the tedium and increase the productivity of the apparatus by automation. Some methods of PVT determination, such as magnetic densimetry [1, 2], lend themselves to automation better than others. The Burnett method [3] of PVT determination appears, at first sight, to be in the class of experiments that are hard to automate. In this method, volume measurement is avoided entirely. Only pressures and temperatures are determined. Pressures are measured along isothermal paths, while repeated expansions are made from a pressurized vessel of volume $V_I$ into an evacuated vessel of volume $V_{II}$. The volume ratio $(V_I + V_{II})/V_I$ is called the cell constant $N$. The pressure ratio $P_r/P_{r+1}$, where $P_r$ denotes the pressure before the $r$-th expansion, will approach the value $N$ for low enough pressures. The initial fill density of a
run, or sequence of expansions, is called the run constant $p_0$. The goal of the experiment is to extract from the sequence of measured pressures the cell constant, the run constant and as many virial coefficients as are required to describe the equation of state of the fluid over the range of densities up to $p_0$ along each isotherm of interest. For nearly ideal gases the number of adjustable parameters is small and the virial coefficients can be determined with great precision [4, 5]. When the fluid is highly nonideal, such as in a near- or supercritical state, many more adjustable parameters are required and loss in accuracy is unavoidable [6]. Obtaining the leading adjustable parameter $N$ in a separate experiment by means of helium expansions is a key to retaining accuracy in the many other adjustable parameters for a highly non-ideal gas. Burnett isotherms cannot cross two-phase boundaries, and therefore the liquid state is not accessible.

The Burnett procedure, with its repeated openings and closings of valves and abrupt changes in the pressure, does not lend itself easily to automation. We have chosen to determine the pressure-density relation by means of the Burnett method along one reference isotherm only. Isochoric (constant volume) measurements, with a pressure measurement at the reference isotherm included for the sake of density determination, can then be used to span the entire range of physical variables. The advantages are the following: (1) The laborious Burnett procedure is limited to one isotherm only. (2) By choosing the reference isotherm in the supercritical regime, the liquid region becomes accessible. (3) It is also possible to enter the two-phase region and determine dew- and bubble point of fluid mixtures.

The automated measurement of pressure entails a certain loss of accuracy because the gages to be used are not as reproducible as the high-quality manually operated piston gages employed for the measurement of the reference Burnett isotherm. The principal cumulative loss of accuracy in highly non-ideal fluids, however, arises from the analysis of the Burnett isotherm itself. Once the reference isotherm is established, not much further loss of accuracy occurs when the isochores are obtained. Moreover, at high, liquid-like densities, a fair amount of uncertainty in the pressure can be tolerated because the density does not change strongly with pressure. The idea of Burnett-isochorically coupled measurements was advocated by Burnett himself [7] and practised by several investigators [6, 8]. In the work reported here we have not fully exploited all beneficial aspects of Burnett-isochoric coupling [8]. Only practical reasons prevented us from doing so at this time.

**EXPERIMENTAL LAYOUT**

The essential components of the semi-automated Burnett-isochoric facility are those outlined in Fig. 1. The heavy-walled nickel Burnett vessels [9], separated by an expansion valve, are connected to a highly sensitive and reproducible diaphragm pressure transducer [10]. A fill valve is used for filling the two vessels and for evacuating the expansion vessel. Transducer and Burnett vessels are immersed in a precision-controlled oil-filled thermostat. The pressure of the gas on the counterbalancing side of the transducer is measured by means of a high-quality gas-operated piston gage. In addition, an accurate measurement of the barometric pressure needs to be performed. During a Burnett run, it is essential that no leaks occur across either valve. The torque used to seal the valves is carefully monitored and increased only incrementally if indications of leakage are found. It is also very important that adsorption on the surfaces of vessels and valves be minimized. Our vessels, with a total volume of approximately 50 cm$^3$, have a
high-quality internal surface finish. In addition, the internal surfaces of the cells, transducer and valves are gold-plated. On the sample side of the system, the gas manifold contains a sample cylinder that can be heated to supercritical temperatures in an oven so as to guarantee homogeneity of the sample in case of a mixture. It also contains the fill and evacuation lines (Fig. 1). On the balancing side, the manifold contains the argon gas supply, various high quality Ruska® gas-operated piston gages and a Ruska® quartz spiral Bourdon gage with digital readout; and an automated gas injector that responds to the offset of the transducer membrane and serves to keep the transducer at null (Fig. 1).

AUTOMATED PRESSURE CONTROL AND MEASUREMENT

The differential pressure transducer designed by Waxman [10] consists of a 2.5 cm diameter, 25 μm thick nickel sheet clamped between two heavy slightly concave back-up surfaces (Fig. 1). The diaphragm position is sensed by means of a capacitance plate imbedded in the top back-up surface. The transducer capacitance is part of a grounded capacitance bridge (Fig. 2) that is located directly above the thermostat. The bridge includes two ceramic capacitors $C_1$, $C_2$, with high temperature stability, and a variable air capacitor $C_3$ that can be manually trimmed to permit rough balancing of the bridge. The bridge voltage is amplified by a differential preamplifier in order to overcome the problems caused by the ground connection, namely the diaphragm. Bridge leads and components have been kept short and rugged so as to minimize fluctuations in stray capacitances to ground. A 30 kHz lock-in amplifier is used as a null detector. The transducer null is the diaphragm position corresponding to equal pressure or both sides. The bridge reading at transducer null is a function of temperature and pressure, because of dimensional changes of the bridge components, and because the dielectric constant of the balancing gas, argon, varies with pressure and temperature. A voltage-controlled variable capacitor $C_v$ (a varicap) is used to set the bridge voltage to zero at transducer null. Preceding a run, the voltage required for the varicap to zero the bridge is calibrated, by interconnecting the two sides of the transducer and varying

Figure 1. Schematic of the Burnett apparatus with filling system, and of pressure injector and pressure gages in the balancing system. Dashed lines indicate heaters. After Ref. 16.
temperature and pressure over the range of interest. The inverse of the varicap voltage required to zero the bridge is found to be a linear function of temperature and pressure. The coefficients are determined by linear least-squares regression. The calibration function is stored in the computer and used to set the varicap voltage at any pressure and temperature in the subsequent run. Over the range of 20\(^\circ\)C and 0 - 10 MPa, the total change of the varicap is about 5 pF; a range of 2-4 V is used for the varicap voltage. The voltage on the varicap can be set manually; in automated operation, however, the microcomputer and a D/A converter provide the voltage (Fig. 2). The reproducibility of the varicap voltage on pressure cycling is better than on temperature cycling; in the latter case, the voltage zeroing the bridge at transducer null varies by at most 0.2 V. This corresponds with a pressure uncertainty of no more than 20 Pa (0.2 mbar), or a corresponding volume uncertainty of \(1.5 \times 10^{-4}\) cm\(^3\). In his paper [10], Waxman claimed a reproducibility of 0.1 mbar for his transducer under isothermal conditions. Thus, no serious deterioration of the instrument has resulted from replacing the original high-quality manually operated variable capacitor by the varicap, and from using the instrument in modes other than the isothermal one.

The automated pressure regulator, described in detail in Refs. 11, 12, is a piston-type gas injector driven by a stepping motor (Fig. 3). Four position switches that are operated by the edge of a metal disk attached to the moving piston determine the mode in which the instrument is operating. In the control mode, when at least one of the center switches is touched, the stepping motor responds to the amplified analog signal from the capacitance bridge of the transducer. When the piston reaches the first microswitch at either end of its travel, a fill or vent valve is opened by a logic system and the piston travels back to the center while
Figure 3. The pressure injector, the "pressure" and "leak" valves operated by its control logic, and the top of the pressure transducer are all connected to the balancing-gas system. The transducer capacitance bridge output drives the stepping motor on the gas injector through an analog signal conditioner and V/F converter. Heavy lines are pressure lines, light ones symbolize electric connections. After Ref. 16.

the transducer is maintained in rough balance. For details, see Ref. 12. Two limiting microswitches serve to switch the injector off in case one of the outer position switches fails. Mechanical features of the injector such as a minimum-backlash ball screw and a high-reduction harmonic drive, make it a highly reproducible and high-resolution instrument. One step of the stepping motor corresponds to 0.25 μm displacement of the piston, or a change of $2.5 \times 10^{-5}$ cm$^3$ in volume, well below the resolution of the transducer. The injector has worked unattended over long periods of time (nights and weekends) without failure.

In the automated isochoric mode, the pressure is presently measured on a quartz Bourdon gage (Ruska Instruments) which has a DC voltage output proportional to the pressure. Our gage has a range of 0 - 10 MPa, 0 - 10 V, is sensitive to 0.1 mV and is claimed to be accurate to 0.01% of full scale, or 1 kPa. The voltage is read on a 5 1/2-digit digital voltmeter with a noise level of ±2 in the least significant digit. By averaging 1000 readings, the noise is reduced to the level of ±5 x $10^{-2}$ mV. We use the instrument in the absolute mode, that is, referring to vacuum in the casing surrounding the quartz spiral, so that we can dispense with barometric readings in the automated mode. The key to optimum performance of the Bourdon gage is keeping track of its zeropoint. We check this zeropoint regularly by evacuating both sides of the gage. We find that the zeropoint drifts slowly but regularly in time, and we refer all pressure values to the properly interpolated zero. We have calibrated the gage with respect to a 3.5 MPa and a 18 MPa range gas-operated piston gage. Our 3.5 MPa piston gage was calibrated at NBS, while the 18 MPa gage was factory-calibrated traceable to NBS. The voltage-pressure calibration of the Bourdon gage was fitted by a linear relation. The standard deviation of the calibration curve was 100 Pa (1 mbar), an order better than claimed by the manufacturer. During the course of several months of data-taking, the readings of the three pressure gages were compared frequently. They never differed by more than 300 Pa.
The Burnett apparatus and transducer are immersed in a circulating silicone-oil bath [9]. A 4-lead platinum resistance element, placed in an AC inductance bridge, serves as a temperature sensor. Two arms of the bridge are formed by a 6-digit programmable ratio transformer. The bridge is operated by a 300 Hz lock-in amplifier, the output of which is conditioned and fed into a programmable power supply which provides the heater current. The short-term stability of the system is about \( \pm 1 \) mK, and isothermal runs stretching over several days had a temperature stability of \( \pm 2 \) mK. The four most significant decades of the ratio transformer are computer-interfaced by means of a set of sixteen programmable switches on an actuator card in the data acquisition system. The temperature can be selected in steps of 0.05 K to assume any value in the range of 273-473 K.

Five-junction thermopiles spaced throughout the bath reveal that thermal gradients are no larger than \( \pm 2 \) mK. Additional small guard heaters are mounted on the tubes entering the top of the bath. A simple controller with proportional and reset action drives a power unit that supplies current to the guard heaters, thus minimizing temperature gradients near the top of the bath.

The temperature is measured by a standard 25 ohm platinum resistance thermometer, inserted in the Burnett apparatus. The resistance is read directly by an 6½-digit computer-interfaced digital voltmeter (DVM), the least-significant digit corresponding to \( \pm 1 \) mK and with accuracy better than \( \pm 2 \) mK. The DVM has some special features such as elimination of thermal EMF's and decreased noise and reduction of the effect of temperature fluctuations by averaging of the readings. Initially, the thermometer was read alternately on the DVM and also on a 7-decade Mueller bridge and an automated AC bridge; the latter two devices each had a resolution of 0.1 mK. Invariably, all temperature readings agreed with each other within 2 mK, even at elevated temperatures, so that, at later stages, only the DVM was read.

AUTOMATED ISOCHORIC MEASUREMENTS

In the isochoric mode, the temperature of the system is varied in steps and, after equilibration, the pressure and temperature are recorded. The temperature settings are targeted ahead of time, and temporarily stored on a floppy disk memory. The temperature is changed by ramping the ratio transformer setting in the AC bridge. The pressure controller tracks the sample pressure by modifying the pressure of the balancing gas so as to keep the transducer nulled. When the ratio transformer reaches a setting corresponding to the desired temperature, the system enters an equilibration period, during which temperature and pressure are read repeatedly. The average of a sample of 30 readings, continuously updated, and its standard deviation are calculated, and the system is considered equilibrated if the standard deviation is less than a preset limit. Once this point is reached, the final temperature and pressure readings are taken and stored on the data disk. The system then proceeds to the next point, etcetera. After temperature changes of a few degrees, typical equilibration times are of the order of an hour.

A WORKING FLUID FOR A BINARY GEOTHERMAL POWER CYCLE

Conventionally, geothermal heat is utilized by expanding steam, flashed off from the hot brine, in a turbine. The disadvantages of this procedure are twofold.
The impurities in the geothermal steam may lead to deposits and corrosion in various parts of the cycle. The relatively low temperatures, typically below 500 K, of the geothermal sources are not optimal for steam power cycles that typically run near to or above the critical point of steam (647 K). It has therefore been proposed to use a working fluid other than steam, with a more suitable critical temperature. A prototype binary geothermal power plant, constructed in Heber, CA, uses a mixture of 90 mol% isobutane, 10 mol% isopentane, as a working fluid, with a critical temperature of 414 K. The mixture runs through an in-part supercritical Rankine cycle. For several years NBS has had a project for determining the thermodynamic surface of isobutane-isopentane mixtures. For details and results, we refer to the literature [13–15]. On the basis of available data for pure isobutane, pure isopentane, the mixture critical line and vapor-liquid equilibrium below 360 K, a preliminary surface was constructed by means of the principle of generalized corresponding states [13–15]. The Burnett-isochoric apparatus was used to obtain PVT data on the 90/10 mixture at temperatures above 360 K and extending into the supercritical regime. These data were then used to fine-tune [14] the original surface [13].

EXPERIMENTAL PROCEDURE

A supercritical reference isotherm at 425 K, was selected from which isochoric runs branch off, which may, if desired, enter the two phase regime. Along the reference isotherm, the cell constant N was determined by a sequence of expansions with helium. Next, several Burnett runs were made with the mixture as the sample. At various points in the course of this work, isochoric excursions were made to lower temperatures. We measured five isochorics, one at liquid, one at near-critical and three at vapor densities. The mixture was prepared by weighing, after volatiles were removed from both components by repeated pumping and freezing. During filling of the apparatus, phase separation had to be avoided at all cost, since it would affect sample composition. We built an oven to house the sample cell (Fig. 1) and heated the cell, all fill lines and valves to well above 425 K. Likewise, Burnett expansions along the reference isotherm had to be carried out with all due care, since rapid expansion from a state of near-critical density might cool the gas into the two-phase region. Once a composition difference establishes itself between the two Burnett cells, the run is lost. We performed the expansions quite slowly and kept temperature drops of the cell body within a few tenth K. We have found that the fill and expansion valves deteriorate rapidly at temperatures above 425 K, even with careful monitoring of the torque. If small leaks develop across the valves, the pressure in the primary Burnett chamber is not significantly affected as long as the pressure gradient across the expansion valve is small. Therefore, we kept the fill and expansion valves both closed during the isochoric runs, with equal pressure in both chambers. This practice has the additional advantage of permitting us to enter the two-phase region on the isochoric runs. With the valve closed, the only composition differences that arise are between primary Burnett vessel and transducer. The very small volume of the transducer appears to work to our advantage, since we obtain reproducible data before and after entering a two-phase region, be it that equilibration times may increase to several hours.

APPARATUS PERFORMANCE

Details of the cell constant determinations and the Burnett reference isotherm at 425 K are given elsewhere [14]. Here, we will discuss the results obtained in the automated isochoric mode. As typical cases we discuss the liquid-density
isochore at 5.86 mol/dm³ and a vapor isochore at 0.58 mol/dm³. In Figs. 4 and 5, we

Figure 4. Percentage departures $\delta P = (P_{\text{exp}} - P_{\text{calc}})/P_{\text{calc}}$ of the 5.86 mol/dm³ isochoric pressures from the thermodynamic surface. The scatter of the data corresponds to less than 300 Pa. o one-phase, x two-phase data.

Figure 5. Percentage departure $\delta P = (P_{\text{exp}} - P_{\text{calc}})/P_{\text{calc}}$ of the 0.58 mol/dm³ isochoric pressures from the thermodynamic surface. In general, the data scatter no more than 100 Pa.

show the departures of the pressures from the optimized thermodynamic surface, with the scale expanded sufficiently to reveal the scatter of the data.

On the liquid isochore, Fig. 4, the scatter of the pressure data (temperature and density assumed to be exact) is no larger than a few 0.01%, less than 300 Pa for this isochore [16]. This is close to the long-term reproducibility of the Bourdon gage. The pressures before and after entering the two-phase region agreed on this level and showed no systematic differences. In Fig. 5, for the low-density vapor isochore at 0.58 mol/dm³, we note a similar scatter of a few 0.01% in the pressure. At these low pressures, the scatter corresponds to only about 100 Pa [16]. Here, pressures after exiting from the two-phase region were about 20 Pa (0.2 mbar) low. This effect is, however, no larger than the irreproducibility of the transducer null on temperature cycling. When the two-phase region is entered, the P-T relation along an isochore undergoes a sharp change in slope. Figs. 6 and 7 display this for
Figure 6. The determination of the bubble point on the 5.86 mol/dm$^3$ liquid isochore. Pressure differences from a low-degree P(T) polynomial fitted to the one-phase data are displayed. A sudden change in slope occurs when the two-phase region is entered.

Figure 7. The determination of the dew point on the 0.58 mol/dm$^3$ vapor isochore. The representation is the same as in Fig. 6.

we have demonstrated that the automated isochoric mode of the Burnett apparatus allows us to take data on a near- and supercritical mixture in the one- and
two-phase regions with a precision to be anticipated from the reproducibility of the quartz Bourdon pressure gage corrected for drift of its zero. In our case, this precision is about 200 Pa, or 0.01% in pressure at a pressure of 2 MPa. Pressures on the reference Burnett isotherm, however, are measured with an accuracy of better than 1 part in 10,000 throughout, and a precision of roughly an order higher. We do not think that the degradation is serious, since the reliability of the isochoric data is determined not only by the pressure measurement, but also by the tolerance of the density value assigned to the isochore. In our experience, the uncertainty in the density assignment exceeds that of the pressure measurement along the isochore. The reason is that, even with exquisite precision of the pressure measurement along the reference isotherm, a serious loss of accuracy occurs in the multiparameter non-linear regression required to establish the pressure-density relation, especially when the gas is nonideal. We estimate the ultimate reliability of the density at fixed pressure to be from 0.05% at pressures up to 1.7 MPa to 0.07% at pressures up to 5 MPa for this particular mixture [17]. At liquid densities, the result of the pressure uncertainty of the Bourdon gage is negligible, given the fact that the liquid density does not depend strongly on pressure. Only near a critical point, or where the pressure is very low, the density error resulting from the use of the Bourdon gage may exceed that resulting from the density uncertainty on the reference isotherm. Nevertheless, it would be worth our while to increase the accuracy of the Bourdon gage at low pressures, since this would permit exploiting other benefits of the Burnett-isochoric coupling method [8]. For instance, performing a complete grid of Burnett-isochoric measurements in the one-phase vapor regime, each of the isothermal data sets represents a Burnett run even if no actual expansions were done at that temperature. That means, that an independent Burnett analysis can be done at each temperature, thus permitting a cross-check on the isochoric results and a test on the absence of adsorption [8]. In that application, however, the precision of the pressure measurement cannot be allowed to deteriorate. The use of a quartz spiral Bourdon gage with a maximum range of 1 MPa instead of 10 MPa, as we are using now, would come a long way towards achieving this goal. We are in the process of acquiring piezoelectric quartz pressure gages as another alternative. Again, however, more than one gage will be needed to span the entire pressure range with sufficient accuracy. These gages will also enable us to expand the range of automated operation to higher pressures.

The only other weak spot of the apparatus is the short life of the valves at high temperatures. Replacing the valves not only requires dismantling of the apparatus, but also subsequent redetermination of the cell constant, and is therefore to be avoided as much as possible. We have tried several types of valves from reputable companies; these valves all performed splendidly below 373 K, but deteriorated above 425 K, requiring rapidly increasing torque to seal. We have just replaced the valves again, with ones that have served up to 550 K to the satisfaction of some of our more demanding colleagues. It is too early to report on the success of this modification.

ACKNOWLEDGMENTS

The apparatus development was supported by the Office of Basic Energy Sciences, Contract No. DOE-EA-77-A-01-6010, Task No. 134, U.S. Department of Energy. Meyer Waxman initiated the automation of the Burnett apparatus by designing and building the automated pressure injector prior to his death, July 13, 1983. At various stages of the work, we have received help and advice from NBS colleagues in our own and in the Pressure and Vacuum Division, while Dr. Richard Davis from the Length and Mass Division made his facilities available for sample preparation.
REFERENCES


*In order to describe materials and experimental procedures adequately, it was occasionally necessary to identify commercial products by manufacturer's name or label. In no instance does such identification imply endorsement by the National Bureau of Standards, nor does it imply that the particular product or equipment is necessarily the best available for the purpose.
PROPERTIES OF ELASTOMERIC LAMINATES

S. K. Clark and R. N. Dodge
Department of Mechanical Engineering and Applied Mechanics
University of Michigan
Ann Arbor, Michigan 48109, U.S.A.

ABSTRACT

Significant energy loss occurs in reinforced elastomers due to internal hysteresis, and most such materials exhibit strong non-linearities both in their real and imaginary response to cyclic loads. Application of FEM techniques to design for minimization of hysteretic loss requires realistic material models which represent laminate response in the low strain region, where most applications reside. The objective of this present paper is to propose a bilinear material property description suitable for most cord-rubber laminates, and to present typical data for commonly used reinforcing materials.

INTRODUCTION

Computer based structural analysis in the form of finite element codes has been highly successful in refining structural design in both isotropic materials and rigid composites. This has led the rubber industry to attempt to make use of such techniques in the design of structural cord-rubber composites, among the more prominent being tires of all types where volume production or technical necessity can justify extensive design computation.

While such efforts appear promising, they have not been easy to achieve for several reasons. Among these is a distinct lack of a clearly defined set of material property descriptors suitable for computer analysis. There are substantial differences between conventional steel, aluminum or even rigid composites such as graphite-epoxy, and textile-cord reinforced rubber. These differences are both conceptual and practical.

RIGID COMPOSITES

Rigid composites usually utilize relatively small diameter filamentary reinforcement with quite high volume fractions, 0.5 to 0.7 being common, so that the concept of smeared or averaged elastic properties is widely accepted as valid. The notation of Fig. 1 allows anisotropic properties of a plane lamina to be expressed in the form of the five elastic constants \( E_1, E_2, G_{12}, \nu_{12}, \) and \( \nu_{21} \), with \( \nu_{12} \) being related to \( \nu_{21} \) through symmetry relations, so that four elastic properties suffice for experimental description. These are frequently taken to be constants although they need not always be so. In many applications the elastic properties are quite close in numerical value for filaments in either tension or compression and so little or no differentiation is made between them.

Finally, it should be noted that while the individual filaments usually are themselves orthotropic, with different properties in the filament direction compared to the direction perpendicular to the filament, they are solid monofilaments and they
are almost always individually embedded in the matrix. They cannot slip or move relative to one another, and so individually and collectively contribute to the material properties of the composite sheet.

ELASTOMERIC COMPOSITES

One major difference between cord-rubber and rigid composites lies in the ratio of reinforcement modulus to matrix modulus. One measure of this is the ratio of modulus of the lamina of Fig. 1 in the cord direction to the transverse direction, $E_1/E_2$.

![Figure 1. Principal Material Directions for Unidirectionally Reinforced Lamina.](image)

![Figure 2. End View of Cord-Rubber Composite](image)

Ref. [1] lists values shown in Table 1 below.

<table>
<thead>
<tr>
<th>Composite System</th>
<th>$E_1/E_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass-epoxy</td>
<td>2.9</td>
</tr>
<tr>
<td>Graphite-epoxy</td>
<td>40</td>
</tr>
<tr>
<td>Nylon-rubber</td>
<td>45</td>
</tr>
<tr>
<td>Polyester-rubber</td>
<td>75</td>
</tr>
<tr>
<td>Fiberglass-rubber</td>
<td>290</td>
</tr>
<tr>
<td>St.-el-rubber</td>
<td>850</td>
</tr>
</tbody>
</table>

Table 1: Ratio of Longitudinal to Transverse Moduli

Practically speaking, cord-rubber composites are characterized by very low modulus matrices which provide very little lateral reinforcement against filament buckling. Further, since the internal section of most strands of filaments are not fully penetrated by the dip, they are supported laterally only by adjacent filaments. Due to twist, the initial shape of individual filaments is curved, and the filaments straighten under tension and become stiffer. All of these effects combine to result in the stress-strain curves of lamina in the one direction being distinctly non-linear for all commercially important cord-rubber laminates. This is illustrated conceptually in Fig. 3, where it is understood that the transition region may occur at different levels of cord strain $\varepsilon_1$ depending on the reinforcing cord material.

Stress-strain response such as illustrated in Fig. 3 has been widely discussed in the literature as bimodular response. Two excellent review articles covering the mathematical description [2] and the mechanics of the phenomenon [3] are available.

Numbers in parentheses refer to references in the Bibliography.
Past work on this phenomenon has divided this non-linear response into regions of cord tension and cord compression. Experimental data shows that this is an oversimplification. Figure 4 shows data on rayon from [4], Fig. 5 shows data on nylon from [5], Fig. 6 shows data on steel-rubber from [6], Fig. 7 shows data on polyester-rubber from [7] while Fig. 8 shows data on aramid-rubber, also from [7].

In view of the clearly different behavior exhibited by these various reinforcing materials, it is proposed that the bimodular simplification be adapted in the form illustrated in Fig. 9, taken from Ref. [5]. Here the stress-strain curve in the reinforcement direction is idealized into two linear regions, I and II. Region I is the low cord modulus region, with individual filaments contributing little to lamina stiffness. In region II the filaments are fully effective in enhancing stiffness. The two regions are linearized and their intersection is expressed in terms of cord strain $\varepsilon^*_1$, a material property of the lamina. The elastic constants of the lamina must be
Figure 6. Stress-Strain Curve for Steel Reinforced Rubber, 1 direction, Ref. [6]

Figure 7. Stress-Strain Curve for Polyester Reinforced Rubber, 1 direction, Ref. [7]

Figure 8. Stress-Strain for Aramid Reinforced Rubber, 1 direction, Ref. [6]

Figure 9. Generalized Stress-Strain Curve for Bimodular Material, with Notation
measured separately in both regions I and II.

In this descriptive framework, an elastic cord-rubber lamina is now described by eleven elastic constants, these being shown below in Table 2.

<table>
<thead>
<tr>
<th>Region I</th>
<th>Region II</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_1)</td>
<td>(E_1)</td>
</tr>
<tr>
<td>(E_2)</td>
<td>(E_2)</td>
</tr>
<tr>
<td>(G_{12})</td>
<td>(G_{12})</td>
</tr>
<tr>
<td>(\nu_{12})</td>
<td>(\nu_{12})</td>
</tr>
<tr>
<td>(\nu_{21})</td>
<td>(\nu_{21})</td>
</tr>
<tr>
<td>plus (\epsilon_1^*)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Elastic Constants for a Cord-Rubber Lamina

In future notation, the elastic constants will be subscripted with I or II to indicate the region in question.

Some elastic constants have been reported in the literature for both regions I and II, such as in refs. [6] and [8]. Constants are given for nylon, fiberglass, and steel from our current studies in Table 3 below.

<table>
<thead>
<tr>
<th>Material</th>
<th>(E_1)</th>
<th>(E_2)</th>
<th>(G_{12})</th>
<th>(\nu_{12})</th>
<th>(\nu_{21})</th>
<th>Vol. Frac.</th>
<th>Cord</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nylon Rubber</td>
<td>88</td>
<td>9.6</td>
<td>3.0</td>
<td>.483</td>
<td>.097</td>
<td>.15</td>
<td>Present Work</td>
</tr>
<tr>
<td>Polyester Rubber</td>
<td>442</td>
<td>9.7</td>
<td>3.7</td>
<td>.431</td>
<td>.011</td>
<td>.18</td>
<td>Present Work</td>
</tr>
<tr>
<td>Fiberglass Rubber</td>
<td>392</td>
<td>10.5</td>
<td>3.8</td>
<td>.430</td>
<td>.233</td>
<td>Present Work</td>
<td></td>
</tr>
<tr>
<td>Steel Rubber</td>
<td>763</td>
<td>10.7</td>
<td>3.9</td>
<td>.402</td>
<td>.0033</td>
<td>.15</td>
<td>Present Work</td>
</tr>
</tbody>
</table>

Table 3: Bimodular Elastic Properties, Modulus Values in Mpa

Insofar as is known, no data has been published on \(\epsilon_1^*\). A review of available \(\sigma_0\) vs \(\epsilon_1\) curves available in the literature and from work done by the present authors gives values of \(\epsilon_1^*\) as shown in Table 4. These should be taken as representative only, since they are clearly a function of variables such as cord twist, compound modulus, and construction of the lamina.

<table>
<thead>
<tr>
<th>Material</th>
<th>(\epsilon_1^*)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nylon 840/2</td>
<td>.015</td>
<td>Present work</td>
</tr>
<tr>
<td>Polyester</td>
<td>0</td>
<td>[7]</td>
</tr>
<tr>
<td>Polyester</td>
<td>.001</td>
<td>Present work</td>
</tr>
<tr>
<td>Fiberglass</td>
<td>-.0005</td>
<td>Present work</td>
</tr>
<tr>
<td>Steel</td>
<td>.0065</td>
<td>[6]</td>
</tr>
<tr>
<td>Steel</td>
<td>.001</td>
<td>Present work</td>
</tr>
<tr>
<td>Aramid</td>
<td>0</td>
<td>[6]</td>
</tr>
</tbody>
</table>

Table 4: Values of \(\epsilon_1^*\) for Various Reinforcement Systems
What is clear from Table 4 is that allowance should be made in general for the bimodular transition to occur at a value of $\varepsilon_1^*$ which is an inherent material property, just as clearly as the other elastic constants. Computationally, this quantity can be tracked and used as a switch to move from one set of values to another. Such a system probably provides a realistic material framework compatible with present FEM technology, as evidence in its use in Ref. [8].

A second major difference between conventional rigid composites and cord-rubber composites lies in the significant energy loss component associated with cyclic stress in cord-rubber laminates. This effect is large enough so that it must be taken into account in describing tests where self-heating can occur. In this paper the notation of Ref. [9] is adapted, so that elastic constants can be expressed as:

$$E^* = E' + iE''$$

where the real portion $E'$ represents the in-phase elastic modulus obtained by using the slope of the line O-1 as illustrated in Fig. 10. The imaginary portion $E''$ is obtained from the well known relation

$$E'' = \frac{A}{\pi \varepsilon_0^2}$$

where $A$ and $\varepsilon_0$ are also defined in Fig. 10.

\[ \text{Figure 10. Typical Stress-Strain Curves. Regions I and II, with Notation.} \]

Since the elastic constants representing the interaction of stress with strain now appear in the form of both real and imaginary parts, the number of elastic constants needed for description of the lamina of Fig. 1 now is 17, as shown in Table 5.

In Table 5, the Poisson's ratios $\nu_{12}$ and $\nu_{21}$ represent the interaction of two strain terms. On physical principles, and from experimental observation where only in-phase values have been observed, these constants are assigned only real $\nu$-values.
Table 5: Elastic Constants for Cord-Rubber Composite

<table>
<thead>
<tr>
<th>$E_{1,I}$</th>
<th>$E_{1,II}$</th>
<th>$E_{2,I}$</th>
<th>$E_{2,II}$</th>
<th>$G_{12,I}$</th>
<th>$G_{12,II}$</th>
<th>$V_{12,I}$</th>
<th>$V_{12,II}$</th>
<th>$\nu_{21,I}$</th>
<th>$\nu_{21,II}$</th>
</tr>
</thead>
</table>

**LAMINATE PROPERTIES**

Rigid laminates have long been demonstrated to exhibit elastic stiffness properties as predicted by transformation theory. It was postulated that such transformation theory would hold in both linear regions of a bimodular complex material description as well.

An extensive series of experiments has been conducted to verify this. The reinforcing materials have been nylon, polyester and fiberglass. Modulus data was obtained using several test techniques, the most common being cyclic extension of tubes, but also involving biaxial stress states obtained by the tube inflation and torsion.

The results of these experiments appear to confirm and extend the concept that transformation theory is applicable to the imaginary elastic modulus as well as the real elastic modulus, as might be anticipated from the correspondence principle.

Figure 11 shows comparisons of the real Young's modulus of nylon in both Region I and Region II with calculated values over the range of cord angles 0° to 90°, while similar data is given for the imaginary modulus in Fig. 12. Shear modulus data is given in Figs. 13 and 14. From these comparisons it appears that satisfactory prediction of complex modulus for any angle can be obtained using such transformation.

**REFERENCES**

Figure 11. Variation of Real Young's Moduli $E'$ with Cord Angle for a Nylon-Rubber Bias Ply Tube.

Figure 12. Variation of Imaginary Young's Modulus $E''$ with Cord Angle for a Nylon-Rubber Bias Ply Tube.

Figure 13. Variation of Real Shear Modulus $G'$ with Cord Angle for a Nylon-Rubber Bias Ply Tube.

Figure 14. Variation of Imaginary Shear Modulus $G''$ with Cord Angle for a Nylon-Rubber Bias Ply Tube.
NEW EXPERIMENTS ON THE LANGMUIR FILM BALANCE

Graham Morrison
Thermophysics Division
National Bureau of Standards
Gaithersburg, MD. 20899

Ian L. Pegg
Physics Department Thermophysics Division
Catholic University of America National Bureau of Standards
Washington, D.C. 20064 Gaithersburg, MD. 20899

ABSTRACT

We describe here the design and construction of a wide-range, automated film balance and a set of experiments on a well-documented material to test the instrument during its development. The balance can be operated in the temperature range 5 - 50 °C. It can measure surface tension differences up to 50 mN m⁻¹ with an imprecision of ±0.01 mN m⁻¹. The measurements of the liquid-expanded (LE)-liquid-compressed (LC) and the liquid-vapor transitions in pentadecanoic acid monolayers on water to confirm recent suggestions that long-standing ideas about these two transitions might be in error and that a thorough re-examination of many "classic" measurements will be necessary.

INTRODUCTION

When certain kinds of materials, typically species with a large, non-polar hydrocarbon hydrocarbon part and a localized, polar functional group that can be involved in hydrogen bonding, are deposited on the surface of the water, they will spread to form a film one molecule thick. Although phenomena associated with the formation of monomolecular films have been known from antiquity, it was not until the second half of the 18th century that the first quantitative studies on such materials were performed; in 1771, Benjamin Franklin reported that a teaspoon of oil, when poured on Clapham Pond, spread to cover an area of nearly half an acre [1]. Franklin did not recognize that the thickness of these films was a measure of molecular size. It was not until the end of the 19th century that Lord Rayleigh made that connection [2]. It was during this same period that Frau Agnes Pockels developed the precursors to modern techniques of handling these films by manipulating them with moveable barriers [3].

In its most basic form, the Langmuir film balance [4] consists of a liquid-vapor surface that has been separated into two parts; the barrier separating the parts of the surface can move freely while, at the same time, not allow material constrained to one part move to the other. If one part of the surface is clean and the other is covered with a monolayer, the surface tensions will differ and a net force will act on the barrier, causing it to move in the direction of the surface with the higher surface tension. By applying a force to restrain the motion of the barrier, one can
determine the difference between the surface tensions. An elaboration common to nearly all balances is an additional barrier that allows the area accessible to the monolayer to be changed, as shown in Fig. 1. One thus measures a force proportional to the difference in surface tension as a function of surface concentration of the surfactant. Because plots of these data are so similar in form to pressure-volume plots for ordinary fluids, one often speaks of the states of the monolayer in terms used for ordinary fluids — vapor, liquid, etc. Similarly the difference in the surface tension can be interpreted as the result of a two-dimensional "pressure" exerted by the surfactant; the surface tension difference is often called the "spreading pressure."

Figure 1. A rudimentary Langmuir film balance: (a) the "spreading pressure" sensing barrier; (b) the torsion wire and lever for exerting the restraining force; (c) the barrier for adjusting the monolayer area.

In the section entitled "Design Considerations," we discuss the problems typically encountered in the Langmuir balance and our efforts to circumvent or minimize those problems. In the section entitled "Experimental," we discuss two experiments that were performed as the instrument was being constructed. These experiments show the versatility and the limits of the instrument as it has been designed. We also discuss the results of those experiments and conclude that, despite the long history of experimentation on monomolecular films, there are fundamental features of these materials that have yet to be studied in sufficient detail to clarify their true character.

DESIGN CONSIDERATIONS

There are four kinds of problems that plague the Langmuir film balance. The first is the problem associated with the manipulation of the monolayer. The fine connections between the sensing barrier and the walls of the trough must be robust enough to prevent leakage, yet they must be fine enough not to perturb the force measurement. The contact between the moving barrier and the sides of the trough must also be leak tight. The second problem is the soiling of the reference and monolayer surfaces. Foreign materials can come from the air above the water surface, the material from which the trough is constructed, the water and the solvent that is used to introduce the surfactant onto the water. A third source of error is incomplete humidification of the air above the water surface and temperature gradients in the instrument. Both can produce unexpected fluctuations in the temperature at the

46
from a small Teflon "pontoon". The seal with the sides of the trough are made with waxed nylon threads, with are attached to Teflon blocks that rest on the edges of the trough; the length of the threads is adjusted to make them as nearly semicircular as possible (ca. 1.6 cm). The movement of the barrier is coupled to the sensor through a triangular wire frame; the base of the triangle is slightly bent so that only the apices touch the midrib of the pontoon. In this configuration, the output of the sensor does not depend upon the level of the water in the trough. The top of the triangular frame is attached by a stiff wire to a piece of 0.007 cm stainless steel foil, thus forming the arm of a cantilever. The distance from the surface of the water to the brass frame is about 15 cm. When the pontoon moves in response to a difference in surface tension, its motion is restrained by the flexing of the foil cantilever. The extent of the flexing is measured by two semiconductor strain gauges affixed to opposite sides of the foil; the gauges form two arms of an AC bridge, whose output is read by a data acquisition system after having been processed by a lock-in amplifier. This arrangement eliminates the need for elaborate steps to isolate the balance from ambient vibration. Because of the sensitivity of the strain gauges to temperature gradients and light, the entire sensing mechanism except for the wire triangle was enclosed in a double-walled cannister with water circulating between the walls for temperature control. Until the differences in surface tension became as high as the surface tension of water (ca 72 mN m$^{-1}$), the output of the bridge was found to be nearly linear with respect to spreading pressure. The practical range of the pressure sensor was from 0.0 to 50.0 mN m$^{-1}$ with an imprecision of 0.01 mN m$^{-1}$.

Figure 2. The film balance sensing mechanism: (a) the pontoon; (b) the wire triangle; (c) the foil cantilever; (d) the semiconductor strain gauges.

The state of the balance is then determined by two kinds of measurements, the number and direction of pulses of the stepper motor determine the location of the moving barrier -- and the area accessible to the surfactant -- and the output of the AC bridge determines the spreading pressure of the surfactant. Both these pieces of information were collected by a data acquisition system (Keithley DAS) and subsequently processed by a computer (IBM-XT). The film balance could then be operated in several modes, an area calibration mode, a modified raw-data mode (bridge output voltage vs fraction of accessible area swept out), and a spreading pressure vs inverse surfactant concentration. Corrections for drift in the zero set-point of the AC bridge and for the slight movement of the sensing barrier as it is pushed by the surface tension difference.

**EXPERIMENTAL**

In this section, we will describe the procedure for preparing the monolayers, some of the characteristic behavior of these films, and two experiments that were performed to test the film balance during its construction. In the two experiments,
liquid-vapor interface; the system is then not at the expected conditions. The final source of error is the time over which a single group of measurements are made in traditional instruments. These times, which may be as long as several days, not only provide an opportunity for contamination of the surface but also cause the experiment to be difficult and tedious.

The water was held in a shallow Teflon* trough mounted on a heavy brass base, which was constructed from two brass plates that had been soldered together. Two sinuous channels had been milled in one of the plates to allow water to circulate inside the base. The four sides of the trough were milled from 5 cm x 2.5 cm Teflon stock. The pieces were dove-tailed together to form a frame with inner dimensions 15 cm x 50.5 cm. The inside edges of the long sides were milled to form a "knife-edge" on which the moveable barrier would ride. The bottom of the trough was made from thin (1 cm) Teflon sheet; the entire Teflon assembly was bolted to the base. Finally, holes assembly was bolted to the base. The thin bottom insured good thermal contact with the base. Finally, holes were drilled into the Teflon part of the trough to receive a platinum resistance thermometer. Before using the trough, it was thoroughly cleaned by inverting it and allowing steam to wash across it for a week. Kim and Cannell, whose work we shall refer to later in this paper, found it necessary to soak their Teflon trough in water for six months before plasticizer and other surface active materials leached from the Teflon ceased to be a problem [5]; the week of "steam cleaning" proved adequate.

A frame constructed from four pieces of Teflon rested on top of the trough and enclosed the space above the water. Brass plates were affixed to the sides of this frame; their temperature was controlled by circulating water through copper tubes soldered to the plates. The lid that covered the trough was constructed from glass plates with 1/2" PVC spacers between them. Temperature controlled water was circulated through cut away channels in the 1/2" PVC. The lid was in two parts in order to receive the "pressure" transducer. Several 0.8 cm holes were drilled through the glass-PVC composite cover to allow access to the water surface. The water in the trough was then completely surrounded by a temperature controlled surface. The water was controlled to +/- 0.01 K. The temperature was measured throughout the instrument; the variation in temperature from one point to the other was found to be on the same order as the control of the temperature of the circulating water.

The barrier used for adjusting the area accessible to the monolayer was made from a 17.5 cm x 1.3 cm x 0.6 cm piece of Teflon. Small recesses were milled at each end to fit onto the knife-edges in the sides of the trough. The barrier was held firmly against the trough by brass blocks that fit over each end of the barrier. These blocks were mechanically connected to a carriage that rides on a screw thread mounted underneath the brass base. By turning the screw thread, controlled by a stepper motor, the barrier can be moved back and forth along the length of the trough. By keeping account of the number of pulses to activate the motor, one can keep track of the location of the barrier. There are two other features of the barrier that need to be noted. The first is a modification to eliminate leaks of the surfactant. Because Teflon is not wetted by water, a small open channel can develop where the barrier meets the edge of the trough; the surfactant can, of course, leak through this channel. A small piece of glass has been affixed to the front of the barrier just above the notch that fits over the knife-edge of the trough. The water, which will wet the glass, can be "attached" to the glass piece, thus closing the channel. Finally, a small Teflon tab has been attached to the center of the barrier. When the tab is rotated in the "up" position, it serves no purpose; when rotated in the "down" position, it can touch the sensing barrier when the two barriers are close enough. When the barrier is operated in this mode, the variable surface area can be automatically calibrated.

The sensing barrier and sensor, shown in Fig. 2, are the major modifications made to what is otherwise a traditional Langmuir film balance. The barrier is constructed
monolayers of pentadecanoic acid, a commercially available material (Sigma) which was purified further by zone refinement, were prepared depositing a known volume (typically 20-30 μl) of a hexane solution of the acid on the water surface; the water had been acidified to pH2 with distilled reagent grade HCl. The solution spreads on the water; within a short time, the hexane evaporates leaving the pentadecanoic acid monolayer behind. The concentration of the spreading solutions were typically 0.005 mol l⁻¹. An important source of contaminants is the spreading solvent; experiments performed after depositing the pure solvent on the surface demonstrated that the solvent was free of any detectible surface active impurities. In all the experiments, the deposited film was expanded. By expanding the film, any undesired impurities become less concentrated; compressing the film would concentrate impurities.

When the concentration of the surfactant on the water surface is so low that the molecules can move about freely, the monolayer is said to be in the "vapor" phase. As the concentration is raised, many surfactants will condense; in analogy with ordinary fluids, the monolayer is said to be in the "liquid phase. This transition in pentadecanoic acid monolayers has been studied in great detail by two groups, Benedek and Hawkins [6], who studied the transition and critical point on pure water Kim and Cannell [5], who studied this phenomenon on water acidified to pH2. Their motivation in doing this experiment was to determine the character of the equation of state around the critical point, which is sensitive to the dimensionality of the material. For example, the exponent describing the coexistence curve in a three dimensional fluid is 0.325 [7]; in a two dimensional fluid, 1/8 [8]. Both groups found their data to be consistent with an exponent of 1/2, the "classical" or "mean field" value. More recently, however, Pethica et al. [9, 10] and Pallas [11], have repeated the experiment as performed by Kim and Cannell. Although they were able to reproduce the earlier spreading pressures at "liquid-vapor" equilibrium, for example 162 μN m⁻¹ at 24.97 °C, they found that the equilibrium "liquid" and "vapor" molecular areas to be quite different, indeed: 45 Å² molecule⁻¹ vs 145 Å² molecule⁻¹ for the "liquid" phase and 1300 Å² molecule⁻¹ vs 403 Å² molecule⁻¹ for the "vapor" phase. Kim and Cannell found the critical point to be at 26.27 °C, the later study found evidence for two-phase equilibrium above this temperature with no evidence of a critical point even at 40 °C. Pethica et al. suggested that the difference is the result higher levels of impurities in the earlier experiments; Pallas suspected that these impurities were contaminants that leached from Kim and Cannell's Teflon trough.

![Figure 3. The spreading pressure for the "liquid-vapor" transition vs molecular surface area for pentadecanoic on pH2 water at 35 °C.](image)
Figure 3 shows a measurement made on this system at 35°C, well above the critical temperature suggested by Kim and Cannell. Two sets of measurements were spliced together to cover the entire range of interest. The scatter in the data are indicative of the experimental bounds of the instrument. Nonetheless, the data clearly show a broad flat region characteristic of the first-order liquid-vapor transition.

As the liquid phase is compressed, it undergoes a transition into another liquid phase. The origin of this transition is thought to be associated with a change in the hydrocarbon part of the molecule from a random configuration to a "straight chained" conformation. Typically, this transition has manifested itself as a "kink" in the isotherm without the flat region characteristic of a first-order transition. The transition was thought to have an undetermined higher order; the onset of the transition was well-defined but the chain ordering was gradual. Pethica and his coworkers [9, 10, 11] have reported data that refute the traditional thinking about this transition; their data show features characteristic of a simple first-order transition. Even more recently, the results of second-order harmonic light scattering experiments [12] have suggested the existence of well defined coexisting phases, a situation consistent with a first-order transition.

Our measurements on both pentadecanoic acid and hexadecanoic acid support the contention that the transition is indeed first order. We suspect that many of the earlier measurements contained artifacts arising from impurities, temperature gradients, and below saturation humidification. We have found that isotherms that resemble those usually obtained can be produced by reducing the humidity or allowing surface contamination to accumulate. Our measurements are shown in figure 4. Each of these isotherms required 30 minutes to perform rather than an equivalent number of hours. The time factor alone allows us to cut the time-dependent factor of contamination by nearly 100.

![Figure 4](image-url)  
Figure 4. The spreading pressure for the "liquid compressed" to "liquid expanded" transition vs molecular surface area for pentadecanoic acid on pH2 water.

CONCLUSION

We have described here a Langmuir film balance that has been designed with two major improvements. The first is the complete automation of the instrument. This allows the instrument to be operated without being handled in any way by the experimenter. It frees the experimenter from much of the tedium associated with this particular experiment and the decline in objectivity associated with that tedium. The second modification is the sensor, no longer a fragile device that can operate over
a few orders of magnitude in spreading pressure, but a robust device that operates over nearly five orders of magnitude in spreading pressure, a range that could be extended with modification of the sensing device. These improvements have made the rapid examination of phenomena in monolayers a possibility and have dramatically reduced the serious problem of time-dependent surface contamination.

Even while in its development stages, use of this instrument has shown that many phenomena in monolayers need to be carefully reexamined. We have shown that two phenomena, which have been exhaustively studied in the past, the vapor-liquid transition and the LE-LC transition, have much information to reveal.

ACKNOWLEDGEMENT

This work was supported in part by the Department of Energy and by the National Bureau of Standards.

* In order to describe materials and experimental procedures adequately, it is occasionally necessary to identify commercial products by manufacturers' name or label. In no instance does such an identification imply endorsement by the National Bureau of Standards, nor does it imply that the particular product or equipment is necessarily the best available.

REFERENCES

2. LORD RAYLEIGH, Phil. Mag., 48, 337 (1899).
Physicochemical Applications of Supercritical Fluid Chromatography

Thomas J. Bruno
Thermophysics Division
National Bureau of Standards
Boulder, Colorado 80303

Abstract

Supercritical fluid chromatography is a technique which offers many unique advantages and opportunities when implemented as a physicochemical measurement tool. A logical intermediary between gas chromatography (G.C.) and high performance liquid chromatography (H.P.L.C.), supercritical fluid chromatography (S.F.C.) finds a natural place in thermophysics research. In this paper, a brief overview of S.F.C. will be given (with reference to a newly constructed instrument), followed by a general description of the measurements made possible by this instrument. In addition, some preliminary measurements will be presented.

Introduction

The general field of chromatography has recently celebrated its eightieth birthday. Since the original experiments were performed by Tsvet [1], chromatography has enjoyed a growth rate virtually unparalleled in any other scientific discipline. The evolution of chromatography can be divided into four distinct periods, each with its innovations, fads, and failures.

During the early years of the development of chromatography, the technique consisted entirely of liquid column methods [2]. This protocol was directed toward the separation of mixtures of colored materials (such as inks and natural botanical products) using glass columns filled with an adsorbent, and elution using common liquid solvents. The adsorbents included materials such as corn starch, firebrick dust and eventually silica and alumina. Typical examples of solvents were lower molecular weight alcohols and water. The efficiency of these methods was quite low, and the results less than spectacular. The separated components were usually not eluted from the column using the solvent, but rather were left as colored bands on the adsorbent. The most common method of product recovery was to force the adsorbent out of the column onto a plate using a rod. The desired components (on the adsorbent) were then sliced out using a spatula, and removed from the adsorbent in a Soxhlet extractor. This first period in the development of chromatography is the only one to have a definite endpoint; the other three periods, to be described in the following paragraphs, began at different times but are all still in progress.

The use of a gas (or an organic vapor) as the eluent marks the start of the second period, and was such a major advance in chromatography that the effects are still in evidence in laboratories throughout the world [3-7]. Gas chromatography has become the generic name of four basic separation techniques which all make use of a gas as the mobile phase. This carrier transports the sample (or analyte) through the so-called stationary phase, in which the separation actually takes place by physicochemical interactions. The nature of the stationary phase (and the major physical interactions taking place in that phase) serve to describe the type of gas chromatography being used. The stationary phase may consist of a solid adsorbent (gas-solid chromatography GSC), a liquid coated on a solid, inert support (gas-liquid
chromatography, GLC), a liquid coated on an active solid support/adsorbent (gas-layer-
adsorption chromatography, GAC), or a liquid coated on the inner walls of a long
glass or fused quartz capillary (capillary or open tubular column chromatography,
CGC).

It is interesting to note that the third period in the development of chroma-
tography was a modern repetition of the first period. The practical and theoretical
advances made during the gas chromatography (second) period were applied to the class-
cial methods of liquid column chromatography described earlier. This work has been
completed only in the last twenty years and has resulted in the emergence of the
field of high performance liquid chromatography (HPLC) [3,9]. As with gas chromato-
graphy, high performance liquid chromatography is a general name applied to a number
of separate techniques, differing in the stationary phase type employed. In each
case, the mobile phase is a liquid solvent or solvent system. The stationary phase
may consist of a liquid, immiscible with the carrier liquid, on a solid support.
This is the liquid-liquid (LL) technique, and includes the bonded phase and ion pair
modifications which are most useful today. The other techniques are liquid adsorp-
tion chromatography, (making use of high surface area solid particles to effect separ-
ation), size exclusion chromatography (using particles of known, uniform pore size),
and ion-exchange chromatography (using an ion-exchange resin to preferentially re-
tain different ionic species).

The fourth (and most recent) major period in the development of chromatography
is the present era of supercritical fluid chromatography (SFC) [10]. The most obvi-
ous difference between SFC and the more traditional GC and HPLC is the use of the
supercritical fluid mobile phase. A supercritical phase can be understood by consid-
ering the vapor liquid equilibrium of a fluid. As the fluid is heated at constant
total volume, the intensive properties of the vapor and the liquid become more simi-
lar. This is especially pronounced in the critical region. At the critical point, a
coalescence of intensive properties takes place. Further heating produces the so-cal-
led supercritical fluid, which will not have a liquid phase regardless of the applied
pressure.

Some properties of a typical supercritical fluid phase can be seen in Table I.
A list of some useful supercritical carriers, along with some relevant properties, is
provided in Table II [11]. The density of the supercritical fluid is very similar to
that of the liquid. This property explains the greatly enhanced solvent power of the
supercritical fluid with respect to a gas. The viscosity of the supercritical fluid
closely resembles gas viscosity, thus allowing relatively easy mass transfer. The
thermal conductivity of a supercritical phase is also relatively large, as one would
expect from density and viscosity considerations. The diffusivity (self-diffusion
coefficient) of the supercritical fluid is intermediate between gas and liquid values.
This gives the supercritical fluid an advantage of mass transfer over liquid-liquid
extraction processes.

<table>
<thead>
<tr>
<th>TABLE I. Comparison of Representative Fluid Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density, ( \rho )</td>
</tr>
<tr>
<td>g/ml</td>
</tr>
<tr>
<td>Diffusivity, ( D )</td>
</tr>
<tr>
<td>cm²/s</td>
</tr>
<tr>
<td>Dynamic Viscosity, ( \eta )</td>
</tr>
<tr>
<td>g/cm s</td>
</tr>
</tbody>
</table>
TABLE II. Some Useful Carriers for Supercritical Fluid Chromatography

<table>
<thead>
<tr>
<th>Fluid</th>
<th>( T_C (°C) )</th>
<th>( \rho_C (g/mL) )</th>
<th>( P_C (MPa) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon Dioxide</td>
<td>31.3</td>
<td>0.460</td>
<td>7.4</td>
</tr>
<tr>
<td>Ammonia</td>
<td>132.3</td>
<td>0.235</td>
<td>11.3</td>
</tr>
<tr>
<td>Nitrous Oxide</td>
<td>36.5</td>
<td>0.450</td>
<td>7.2</td>
</tr>
<tr>
<td>Sulfur Dioxide</td>
<td>157.5</td>
<td>0.520</td>
<td>7.9</td>
</tr>
<tr>
<td>Water</td>
<td>374.4</td>
<td>0.40</td>
<td>23.0</td>
</tr>
<tr>
<td>Methanol</td>
<td>240.5</td>
<td>0.272</td>
<td>8.0</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>235.3</td>
<td>0.273</td>
<td>4.8</td>
</tr>
<tr>
<td>n-Pentane</td>
<td>196.6</td>
<td>0.232</td>
<td>3.4</td>
</tr>
<tr>
<td>n-Hexane</td>
<td>234.2</td>
<td>0.234</td>
<td>3.0</td>
</tr>
<tr>
<td>Dichlorofluoromethane</td>
<td>178.5</td>
<td>0.522</td>
<td>5.2</td>
</tr>
<tr>
<td>Trichlorofluoromethane</td>
<td>196.6</td>
<td>0.554</td>
<td>4.2</td>
</tr>
<tr>
<td>Chlorotrifluoromethane</td>
<td>28.8</td>
<td>0.578</td>
<td>4.0</td>
</tr>
<tr>
<td>dichlorotetrafluoroethane</td>
<td>146.1</td>
<td>0.582</td>
<td>3.6</td>
</tr>
<tr>
<td>dichlorodifluoromethane</td>
<td>111.7</td>
<td>0.558</td>
<td>4.0</td>
</tr>
<tr>
<td>Benzene</td>
<td>288.9</td>
<td>0.304</td>
<td>4.9</td>
</tr>
<tr>
<td>Xenon</td>
<td>16.6</td>
<td>1.155</td>
<td>5.9</td>
</tr>
<tr>
<td>Toluene</td>
<td>320.8</td>
<td>0.29</td>
<td>4.2</td>
</tr>
</tbody>
</table>

The instrumentation for SFC is very similar to that of modern high performance liquid chromatography. The reason for this is that the critical pressures of many useful SFC mobile phases are between 2 and 20 MPa, comparable to the working pressure ranges of LC instruments. A general schematic for an SFC specifically designed and constructed for physicochemical measurement is shown in Figure 1.

![Figure 1. A Schematic diagram of a supercritical fluid chromatograph.](image)

A Physicochemical Supercritical Fluid Chromatograph

For a carrier fluid delivery system to be suitable for supercritical fluid chromatography (especially for use in the measurement of physicochemical parameters), it must be able to generate and maintain a uniform, supercritical pressure at the desir-
ed flow rate. Pressure programming, analogous to temperature programming in gas chromatography, is also desirable but not essential for physical studies. The main part of the fluid delivery system depicted in Figure 1 is a double headed piston pump (C) which is equipped with an electronic pulse elimination circuit. As indicated in the figure, this delivery system is designed for use with high critical point fluids (supplied to the pump as a liquid, B) or low critical point fluids (supplied to the pump as a high pressure gas, A). The use of a liquid (B) is quite straightforward and is similar to the corresponding operation in HPLC. The use of a gas (A) is complicated somewhat by the difficulties of pumping a gas using low volume piston-check valve pumps. To circumvent this problem, the pump heads are refrigerated using a Ranque-Hilsch vortex tube, to allow the pumping of the fluid in the liquid state. In practice, the gas passes through a chilled heat exchanger immediately before the pump heads, thus insuring the arrival of liquid at the pump. The use of the vortex tube in this application is particularly convenient since the components to be chilled are somewhat large and are of irregular geometry.

Following the exit check valve of the pump, the fluid passes through a pressure transducer and a conditioning coil (D). The conditioning coil serves as a miniature surge tank to smooth out small residual pressure pulsations not removed by the compensation electronics mentioned earlier. A forced air oven, depicted as the dashed line in Figure 1, serves to provide the supercritical temperature. This oven has a maximum temperature of 350°C, and a measured temperature uniformity of 0.03°C at 150°C. A high degree of uniformity is obtained by copper baffling which was installed to insure vigorous mixing of the air space within the oven. This oven is equipped with an inert gas purge and high temperature limit cut off for safety. A 25 cm length of coiled stainless steel tubing (E) located inside the oven serves as a heat exchanger to flash the pumped fluid into a supercritical state. Incorporated as part of this heat exchanger is the sensor of a magnetic densimeter which allows an independent and direct measurement of the initial density of the fluid.

Sample introduction is achieved with a high pressure modification (F) of a gas sampling system developed earlier [15]. A six-port chromatographic sampling valve is the heart of this sampling system, along with an evacuable sample loop and high pressure sample reservoirs. It is essential for a number of reasons that for physicochemical applications, samples should be injected in the absence of a separate solvent. This differs from conventional gas chromatography in which samples should be injected as dissolved in a separate solvent. This differs from conventional gas chromatography in which samples are usually injected in a solution of an appropriate solvent such as acetone or an alcohol. In SFC, it is sometimes possible to inject a sample as a supercritical solution in the carrier fluid (for example, a solution of naphtalene in supercritical carbon dioxide injected in a carrier stream of supercritical carbon dioxide). This becomes more impractical as the solute (sample) molecular weight increases, however. To overcome this problem, the injection of samples adsorbed on glass has been performed using the modified sample loop shown in Figure 2. To use this technique, one applies a solvent-borne sample to the sintered glass inside the sample loop using an ordinary HPLC syringe. Heat supplied by the coil and applied rough vacuum serve to remove the solvent, leaving the non-volatile solute adsorbed on the sintered glass. The supercritical carrier is then allowed to enter the evacuated sample loop and dissolve the solute. After a period of equilibration, the valve position is switched to allow the sample to proceed out of the injector. Valve switching is done using a helium-supplied pneumatic actuator (equipped with pilot valves to enhance the switching speed).
The heart of any chromatographic system is the column (G). In SFC, the column can be a packed bonded phase material (similar to that used in HPLC), an immobilized wall coated capillary, or an uncoated open tube. The capillary columns are excellent for analytical applications because of their high efficiency and low pressure drop. The packed columns are somewhat more problematical due to their inherently lower efficiency and large pressure drops (sometimes reaching as high as 60 atm.). For physicochemical measurements, however, packed columns have found some degree of usefulness, along with uncoated open tubes.

Only recently are coated capillary columns being used for physical studies. The specific uses for each type of column will be treated in greater depth during the discussion of the actual physical measurements in the next section.

Detection of the eluted compounds is done as close to the column exit as possible. The instrument described here employs a modified flame ionization detector (denoted by H in the figure, hydrogen - air flame, 160 volts between base and collector). The supercritical fluid stream is fed into the flame from a 20 μm restrictor. It is at the end of this restrictor (inside the flame) that decompression of the carrier stream actually occurs. This configuration eliminates the possibility of solute condensation before detection, and minimizes extra-column broadening of the recorded peaks. The peaks are recorded on a standard electronic integrator (K). The instrument has provisions for additional detectors, particularly an ultraviolet spectrophotometer (J). This detector is preceded by a decompression area (I) which drops the carrier stream to sub-critical conditions. This arrangement would be of value when using a high critical point fluid such as pentane, because of its high response to the flame ionization detector and lack of a chromophore.

Physicochemical Applications of Supercritical Fluid Chromatography

The application of the chromatographic experiment to the measurement of physicochemical parameters may at first glance seem unusual, since one normally associates chromatography with chemical analysis. When one considers that chromatographic separations are effected by the physical interactions between the molecules of the stationary with those of the solute (or analyte), the logic of measurement applications becomes more clear. Conventional gas chromatography has been used for such studies as the thermodynamics of mixing and association, measurement of activity coefficients, partition coefficients, diffusion coefficients, virial coefficients and a number of more applied engineering parameters [17,18]. We will discuss here a few of the physicochemical measurements which are especially amenable to application in a supercritical fluid carrier.

Diffusion Phenomena

Diffusion can be defined as the mutual permeation of two or more different substances, due to the kinetic activity of their molecules, to produce a uniform mixture. The primary descriptor of diffusion phenomena is the diffusion coefficient $D_{12}$, which can be thought of as a proportionality between the flux of a specific molecular species across an arbitrary plane relative to a plane across which there is no net molal flow. If a substance is injected into a chromatograph as a sharp spike, the spike will broaden as it traverses the length of the tube. This phenomena can be better understood by reference to Figure 3. As it assumes the characteristic para...
bolic flow pattern, the initial plug of sample will spread out and eventually, at
time t, resemble a Gaussian curve. This assumes that no interactions occur between
the sample molecule and the wall of the chromatographic column (i.e., no liquid phase
is present on the inside wall of the column). This is an unattainable ideal situa-
tion, however, since there will always be some degree of interaction with a real sur-
face. Provided that steps are taken to minimize these interactions, the approxima-
tion to a normal curve is justified. Considering that the gaussian can be described
by a variance, following Taylor [19,20] and Aris [21] we may write:

\[ \sigma^2 = \frac{2D_{12}L}{u} + \frac{r^2uL}{24D_{12}} \]  

where \( D \) is the interdiffusion coefficient (of species 1 into species 2), \( L \) is the
length of the tube, \( r \) is the radius of the tube, and \( u \) is the linear velocity of the
solute or sample. A simple relationship exists between this variance and the famil-
liar height equivalent to a theoretical plate (HETP) used in analytical chromatogra-
phy:

\[ H = \frac{\sigma^2}{L} \]  

where \( H \) is the HETP. The value of \( H \) is readily obtainable from experimentally access-
able parameters using:

\[ H = \frac{LW^2}{5.54tR^2} \]  

where \( W^2 \) is the peak width at one half of the peak height (refer to figure 3), and
\( t \) is the residence time of the sample inside the column. Combining equations 1-3, we
obtain:

\[ D_{12} = \frac{L}{4tR} \left[ H + \sqrt{H^2 - r^2/3} \right] \]  

The meaningful root is determined from considerations of the carrier velocity, and
the reader is referred to the work of Grushka for details [22].
An example of this type of measurement can be seen in figure 4. This chromatogram was recorded using the instrument described in the previous section. The sample in this case was decane in supercritical carbon dioxide at 20.7 MPa (3000 psi), and 100°C. The high degree of symmetry indicates the absence of appreciable adsorption on the inside walls of the diffusion tube. By measuring the carrier velocity, peak width at half height, and the retention time, one can then calculate a diffusion constant under temperature and pressure conditions mentioned above. A more comprehensive set of diffusion coefficient measurements on the light hydrocarbons will be presented at a later date.

The Capacity Ratio

The most attractive feature of supercritical fluid technology is the high solubility many substances have in the high density phase. The chromatographic experiment offers a convenient parameter through which to study the dependence of solubility on pressure (or more specifically density), temperature and individual fluid. The capacity ratio (or capacity factor, \( k \)) can be defined as:

\[
k_i = \frac{c_{\text{stat}}}{c_{\text{mob}}} \cdot \frac{V_{\text{stat}}}{V_{\text{mob}}} = K_i \frac{V_{\text{stat}}}{V_{\text{mob}}}
\]

where \( c_{\text{stat}} \) is the concentration of sample in the stationary phase, \( c_{\text{mob}} \) is the concentration of sample in the mobile (supercritical) phase, \( V_{\text{stat}} \) and \( V_{\text{mob}} \) are the volumes of these phases, and \( K_i \) is a type of equilibrium constant called the distribution coefficient. The capacity ratio is easily determined from the chromatographic experiment using:

\[
k_i = \frac{t_r - t_0}{t_0}
\]

where \( t_0 \) is the retention time of an unretained species (i.e., the column void volume). The distribution coefficient has all of the familiar thermodynamic consequences of any equilibrium constant (such as the relationship with enthalpy and entropy). The pressure and temperature dependence of the capacity ratio is very similar to that of the distribution coefficient, and can be used to provide information on that more experimentally inaccessible value. The principal interest in the capacity ratio is, however, its reflection of the solvent power of the supercritical phase. In general, the capacity ratio will decrease with increasing pressure or density. The pressure dependence is necessarily nonlinear because of the shape of the PVT (pressure-volume-temperature) surface of the supercritical fluid near the critical region. The density dependence will have a much more linear appearance and is the variable of choice in determining the optimal solubility conditions. The dependence of the capacity ratio on temperature is not as dramatic as the pressure or density dependence. At constant pressure, an increase in temperature will cause an increase in the capacity ratio (indicative of lower solubility) due to the correspondingly lower density of the supercritical phase.

Measurement of the capacity factor in a chromatographic experiment is a low cost method of determining optimum conditions for separations to be performed on an industrial scale [23]. This is done by determining the marginal decrease in capacity factor obtained for a marginal increase in energy (required to produce the marginally higher density in the separator) in the design of a process.
Partial Molar Volumes

It is in general very difficult to measure partial molar volumes of fluids in mixtures at high pressures. A large amount of accurate PVT data are required of both the pure components and the mixtures. The problem becomes even more difficult when one of the components is present at a low mole fraction, such as is often the case of the solute in supercritical extraction or chromatography. It is possible to estimate partial molar volumes using yet another physicochemical application of SFC [24]. Based upon a model developed for adsorption from solutions, Van Wassen derived a simple relationship between the capacity ratio and the partial molar volume of the solute:

\[
\left( \frac{\partial \ln k_1}{\partial p} \right)_T = \frac{V_i - xV_m}{RT} - \chi
\]

where \( V_i \) is the partial molar volume of the solute at infinite dilution in the mobile phase, \( V_m \) is the molar volume of the pure mobile phase, \( \chi \) is the compressibility of the mobile phase, \( R \) is the ideal gas constant and \( T \) is the thermodynamic temperature. The factor \( x \) is the ratio of the partial molar surface areas of the solvent and the solute, respectively. This equation allows the partial molar volumes to be determined from the slope of ln(k) isotherms. An interesting application of this approach was the determination of the infinite dilution partial molar volumes of naphthalene and fluorene in supercritical carbon dioxide near its critical point.

ACKNOWLEDGEMENTS:

The assistance of T.G. Waldorf with some aspects of instrument fabrication is gratefully acknowledged, as well as the financial support of the United States Department of Energy.

REFERENCES

A HIGH TEMPERATURE THERMAL CONDUCTIVITY APPARATUS FOR FLUIDS

H. M. Roder
National Bureau of Standards
Boulder, Colorado 80303, U.S.A.

ABSTRACT

A new apparatus for measuring both thermal conductivity and thermal diffusivity of fluids at high temperatures is described. The technique employed is that of the transient hot wire. Measurements are made with a 12.7 μm diameter platinum wire at times of up to 1 second. The data acquisition system is controlled by a microcomputer and includes several programmable digital voltmeters. The hot wire and a shorter compensating hot wire are arranged in different arms of a Wheatstone bridge. The cell containing the core of the apparatus is designed to accommodate pressures from near zero to 70 MPa and temperatures from 0 to 500°C. For thermal conductivity, the precision of the new system is expected to be around 0.3% and the accuracy 1.0%. For thermal diffusivity the accuracy is estimated to be around 5%. From the two variables measured, we can obtain values of the specific heat, C_p, of the fluid, provided that the density is either measured, or available through an equation of state.

INTRODUCTION

The transient hot wire method is recognized today as the method of highest accuracy for the measurement of the thermal conductivity of fluids for conditions removed from the critical region proper. For example, data obtained near the saturation lines of toluene and n-heptane with instruments of this type have been proposed recently as standard reference data for the thermal conductivity of liquids near ambient temperatures [1]. At cryogenic temperatures, measurements on liquid methane [2,3] and on liquid argon [4,5], made at two different laboratories with different instruments, agree to within 0.5%; the argon values have also been recommended as standard reference data [5].

The apparatus described here is based on a similar, earlier system [6] but incorporates a number of changes. The most important of these changes are the difference in the experimental temperature range and the ability to measure thermal diffusivity. These and other changes, made to improve both precision and accuracy of the system, are described below. The motivation to measure thermal diffusivity is to obtain values of the specific heat, C_p, of the fluid [7], a variable that, in practical terms, is far more important than either thermal conductivity or thermal diffusivity. Of course, to get specific heat from the measured data requires knowledge of the density, either by direct measurement or from an equation of state for the fluid in question.

METHOD

The transient hot wire system has been defined as an absolute primary instrument [8]. The working equation is based on a specific solution of Fourier's law, which can
be found in standard texts (see, for example, reference [9], page 261),

$$\Delta T = \frac{Q}{4\pi \lambda} \ln \left(\frac{4Kt}{a^2C}\right). \tag{1}$$

In Eq. (1), $Q$ is the applied power, $\lambda$ the thermal conductivity, $K(K = \lambda/\rho C_p)$ the thermal diffusivity, $a$ the wire radius, and $C$ a constant equal to the exponential of Euler's constant. If we expand Eq. (1) to separate time from other variables in the logarithm,

$$\Delta T = \frac{Q}{4\pi \lambda} \ln \left(\frac{4Kt}{a^2C}\right) + \frac{Q}{4\pi \lambda} \ln (t) \quad \text{or} \quad (2a)$$

$$\Delta T = A + B \ln (t), \quad (2b)$$

we see that the working equation is a linear relation between the temperature rise, $\Delta T$, at the wire and the logarithm of the time, $t$, after application of a step voltage to the wire. The actual experimental setup approximates as closely as possible the relation given in Eq. (1) and accounts for the deviations from ideal circumstances through a number of corrections [10]. In our apparatus we measure the temperature rise in the hot wire at 250 fixed times 4 ms apart with a modified Wheatstone bridge, and use linear regression to arrive at the two coefficients of the straight line, $A$ and $B$ in Eq. (2b). Comparing Eqs. (2a) and (2b) we can see that the slope of the line is inversely proportional to the thermal conductivity, i.e., we evaluate the thermal conductivity from

$$\lambda = \frac{Q}{(4\pi B)}. \tag{3}$$

To obtain the thermal diffusivity we solve Eq. (1) for $K$, substituting $B$ for $Q/(4\pi \lambda)$:

$$K = \frac{a^2C}{4t}e^{(\Delta T/B)}. \tag{4}$$

The unknowns in Eq. (4) are $\Delta T$ and $t$. If we evaluate $\Delta T$ at a specific time, Eq. (2b), we can obtain a value of the thermal diffusivity $K$ from Eq. (4). The very simplest solution is to set $t$ equal to 1 second in Eq. (2b), i.e., $\Delta T(1) = A$, and Eq. (4) reduces to [11]:

$$K = \frac{a^2C}{4\cdot 1}e^{(A/B)}. \tag{5}$$

Equation (5) is normally used to evaluate $K$ directly from the two regression coefficients. Finally, the heat capacity is obtained from Eq. (6) using the measured values of thermal conductivity and thermal diffusivity and a density, $\rho$, obtained from an equation of state.

$$C_p = \frac{\lambda}{\rho K}. \tag{6}$$

**APPARATUS**

The apparatus is quite similar to an earlier low temperature system [6]. It includes the following elements: the hot wires, the high pressure cell with the wire supports, the Wheatstone bridge, the oven, the measuring and control circuitry, the sample handling system, and the microcomputer. Changes fall into three categories, those associated with the change in experimental temperature range, those made to
improve the precision and accuracy of the thermal conductivity measurement, and those made to enable the measurement of the thermal diffusivity.

Changes associated with the change in temperature range include a new pressure cell and an oven with new temperature controls (Fig. 1). In addition, the sample handling manifold, which will be colder than the cell, was placed in a position below the cell to avoid sample flow due to gravity. The high pressure cell is commercially available and is rated at 80 MPa (12000 psi) for a temperature of 425°C (800°F). Its cylindrical internal working space is 2.54 cm in diameter by 25 cm in depth. Access to the working space in the cell is through a 1/4 inch tube which contains the wires going into the cell and conducts the sample into the cell. The high pressure seal for the wires is commercially available and accommodates eight leads. It is placed outside of the high temperature environment at ambient temperature.

The oven replaces the cryostat of the low temperature system. A 1/2 inch aluminum shield surrounds the pressure cell. The shield is in turn placed inside cylindrical heaters with top and bottom openings covered by square heating elements. The heating elements are further surrounded in all directions with about 30 cm of additional insulation. Two programable power supplies complete the high temperature environment. The temperature of the cell is monitored by a calibrated platinum resistance thermometer (PRT). The thermal gradient along the cell is measured with two resistance temperature devices (RTD's) which are also used to control the various heaters. All three thermometers are monitored with a high precision digital voltmeter which is connected to the computer by an IEEE-488 bus.

Changes made to improve the precision and accuracy of the thermal conductivity measurement include changes in the hot wires, changes in the Wheatstone bridge and the addition of a specialized digital voltmeter (DVM). The new pressure cell allows a length of 19 cm for the long hot wire. This is nearly twice that of the low temperature system and should improve the accuracy of the temperature rise measurement. The new cell also accommodates eight leads into the cell. With seven leads rather than the three previous, Fig. 2, it is now possible to measure the voltages across both long and short hot wires directly. This eliminates the need to account for lead (nuisance) resistances and their dependence on temperature within the cell. The Wheatstone bridge circuit has been changed to improve the accuracy with which the arms of the bridge, the hot wire resistances, and the balance of the bridge can be measured; see Fig. 3. This was accomplished by adding a digital voltmeter to the system, capable of measuring voltages to 0.5 microvolt at the 50 millivolt level. The voltages required in the wire calibration and bridge balancing cycle are fed to the voltmeter through a new multiplexer. As a result, we can now measure each resistance with an accuracy of 4 milliohms, nearly a factor 10 better than in the earlier instrument, and the bridge balance is accurate to 0.5 microvolts. Each arm of the bridge is about 200 ohms at ambient temperature and includes a series of new precision decade resistances. Because the arms are now larger than in the old system, it is possible to include a calibrated 100 ohm standard resistor in each side of the bridge. Thus, when balancing the bridge, the current in each side of the bridge can be measured independently and the reed switch on the interior of the earlier bridge, of nominal resistance 0.1 ohm, is eliminated. Finally, with the new voltmeter, the voltage applied to the bridge can now be measured directly, eliminating a standard resistor that used to be in series with the bridge and several corrections that had to be applied to the bridge power.

The primary measurements in this experiment are the voltages across the bridge, i.e., how much the bridge is out of balance from the initial setting at each given time. In the earlier apparatus these measurements were made with an A/D converter within the computer which had a noise level of around 25 microvolts for measurements of up to 5 millivolt. In the present apparatus, an instrumentation amplifier with a gain of 100 and a filter, is used with the digital voltmeter, resulting in a noise level of 3 microvolt. The corresponding precision in the temperature rises has improved to 0.2% as inferred from the linear regression statistics. Since the use of a filter results in a certain time lag, the A/D acquisition capability of the computer is still used but
Figure 1. High Pressure Cell, Shield and Heaters.
Changes required to enable the measurement of the thermal diffusivity became evident when we attempted to evaluate the thermal diffusivity from our liquid argon measurements [5]. These results were obtained with the old system several years ago before any of the new circuitry had been installed. We found that the thermal diffusivities derived from the experiment varied by ±20% from those derived from a previous correlation [12] and the equation of state [13]. Looking at Eqs. (4) and (5) we see that the wire radius is important, but the exponential term more so. In the experiment, we try to hold the temperature rise, ΔT, to about 4 K, i.e., the value of A is around 4 K. The corresponding values of the slope, B, are around 0.4 K and they are determined with an accuracy of around 0.5%. If we assume the error in the temperature rise or A to be 2% or about 80 mK, then a few short calculations show the error in diffusivity to be 20%. Roughly then, the error in diffusivity is a factor of 10 larger than the error in A, while the error in B affects the diffusivity very little. In other words it is the error in absolute temperature rise that is most significant.

Analyzing the older experiment for errors of 80 mK in the absolute temperature rise, ΔT, reveals several possibilities. Most important is the balance of the bridge. The bridge is normally balanced with a very small voltage, say 50 mV. However, the voltage applied during an experimental run is much larger, say 5 V. A small initial imbalance, say 1 μV, is amplified by the ratio (applied voltage)/(calibration voltage),
Figure 3. The Wheatstone Bridge Schematic for the High Temperature Apparatus. Potential taps are indicated by points A-L.
a factor of around 100, to 100 μV, which is equivalent to 100 mK. Nearly as important is the accuracy of each of the four bridge arms since these are used in the bridge equation to solve for the resistance change. In the old system the bridge arms were 100 ohm measured to ± 0.03 ohm, an error again equivalent to nearly 0.1 K.

We installed the new circuitry, voltmeters, multiplexers and switches in both the new high temperature system and in the old low temperature apparatus. We tested the new arrangement by measuring a new series of isotherms in the single phase supercritical region of argon in the old apparatus. The temperatures were 172, 200, 220, 275, 300 and 325 K. The variables obtained were thermal conductivity, thermal diffusivity and specific heat. So far these results are unpublished except for some preliminary results on the heat capacity, C₀ [7]. In these measurements the precision of thermal conductivity improved to 0.2% as determined from the statistics of the linear regression fits. The accuracy of the thermal conductivity improved as well, because comparison to thermal conductivities obtained in Lisbon at low densities and a temperature of 172 K [4] shows differences of less than 0.5%. The error in thermal diffusivity decreased to around 5% based on a comparison made through the specific heat at a temperature of 275 K [7]. From these tests we infer that errors for the new high temperature system should be as follows: precision and accuracy of thermal conductivity 0.3% and 1.0% respectively; accuracy of thermal diffusivity and specific heat around 5%.

Other uncertainties in the old version of the old system involved channel bias for the A/D acquisition equipment, a switch of unknown resistance interior to the bridge, and errors in the timing of the experiment. All of these problems were addressed in the changes discussed above except for timing. Improvements in the timing are based on using the clock board of the computer to provide the triggers for both computer A/D and digital voltmeter at the same time. Further, the power switch is activated by the very first and the very last trigger coming from the clockboard. Finally, the delay time for the closing of the switch was measured very accurately with a digital oscilloscope after the entire system had been assembled.

PLANS

All parts of the apparatus have been received and assembled. At the same time, nearly all of the circuitry, the new voltmeters, multiplexers and switches were also installed in the old system. Thus, it was possible to conduct much of the debugging and testing for the high temperature apparatus on the old system. Measurements made with the old system form the basis from which we estimated the errors applicable to the new apparatus. A very limited set of tests on the new system have been carried out with hot wires of 25 μm diameter. A set of the smaller 12.7 μm diameter wires will be mounted next. They will be cycled in temperature to anneal them, and to calibrate them and the RTD's against the PRT. Performance tests measuring thermal conductivity and specific heat of argon and nitrogen are to follow.

Since the original analysis of the theory of the instrument [10] considered only those corrections which affected the determination of the thermal conductivity, a revised analysis is called for which examines the effect of the corrections upon the thermal diffusivity and includes second order corrections where appropriate [14].

REFERENCES


RATCHETTING BEHAVIOR OF TYPE 304 STAINLESS STEEL AT ROOM AND ELEVATED TEMPERATURE

Marina Ruggles and Erhard Krempl
Mechanics of Materials Laboratory
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

ABSTRACT

The zero-to-tension ratchetting behavior was investigated under uniaxial loading at room temperature and at 550, 600 and 650°C. In History I the stress level of ratchetting was equal to the stress reached at one percent strain. For History II the stress level was established to be the stress reached after a 2100 s relaxation at one percent strain. Significant ratchetting was observed for History I at room temperature but not at the elevated temperatures. The accumulated ratchet strain increases with decreasing stress rate. For History II no ratchetting strain was observed at room temperature independent of the stress rates used. This observation is explained in the context of the viscoplasticity theory based on overstress by the exhaustion of the viscous contribution to the stress during relaxation. The viscous part of the stress is the driving force for the ratchetting in History I. Strain aging is presumably responsible for the lack of short-time inelastic deformation resulting in a nearly rate-independent behavior at the elevated temperatures.

INTRODUCTION

The prevention of progressive deformation under cyclic loading and stress boundary conditions is a concern in the design of pressure vessels and has resulted in special design approaches [2-9]. Only within the most recent past has ratchetting attracted the interest of researchers in the area of constitutive equations (see [10] and the references contained therein). Almost all analyses so far use either time independent plasticity theory or plasticity in conjunction with creep. It has been shown [10] that time independent plasticity theory almost always predicts less ratchetting than actually observed, even if advanced theories such as kinematic and multiple yield surface theories are used. The question arises where the origin of these discrepancies may lie.

Over the past several years the second author and his co-workers have shown that the inelastic deformation of engineering alloys can be significantly time dependent [11-13], even at room temperature. Their room temperature experiments suggested that one of the driving forces for ratchetting may be the viscous nature of the inelastic
deformation which is very pronounced for type 304 stainless steel. As a consequence a special test program was devised that would accentuate the role of the viscous deformation. This was accomplished by selecting test conditions which would result in no ratchetting if the material were to behave in a time independent fashion and were to obey the postulates of yield surface theories. The experiments were performed at room temperature and at elevated temperature, at 550, 600 and 650°C. The results show that under the selected test conditions ratchetting is viscous in nature and they suggest that improvements in the predictions can be made with the help of a visco-plasticity theory.

TESTING PROCEDURE

The material tested was AISI Type 304 austenitic stainless steel with the following chemical composition: 0.044% C; 1.267% Mn; 0.033% P; 0.016% S; 0.45% Si; 9.5% Ni; 18.64% Cr; 0.34% Mo; 0.25% Cu, taken from reference heat 9T-2796. Prior to testing, machined specimens were annealed in vacuum at 2000°F for 90 minutes (the shape of the specimen was similar to that shown in Figure 1 of [11]).

A servocontrolled MTS axial-torsion testing machine together with the MTS 463 Data/Control processor was used for computerized testing and data acquisition accomplished by a program written in MTS Basic. Displacement was measured on the gage section with an MTS axial clip-on extensometer. The load (engineering stress), displacement (engineering strain), and the command signal (stress or strain) were measured and recorded digitally and stored on the floppy disks. In all cases XY-plots were taken. The data acquisition intervals were 2.9 MPa except during creep tests. During the creep tests a data point was recorded whenever the strain changed by 0.0075%. After the test, digitized data can be recalled for processing and interpretation. The entire history is available for analysis.

TESTING PROGRAM

Two different test histories were performed and are explained with the help of Figure 1.

OA - strain controlled loading of a virgin specimen to a strain of 1.00% with the strain rate of $8.33 \times 10^{-5}$ s$^{-1}$. At point A stress was recorded, which served as the ratchetting stress level for History I. In case of History II, a 2100 s relaxation test at the strain of 1.00% was performed (Segment AA' in Figure 1) and the stress at the end of the relaxation test was used as the ratchetting stress level for History II. The relaxation test was introduced to eliminate the viscous contribution to the stress as postulated by the viscoplasticity theory based on overstress by Krempl et al. [17]. At the end of the relaxation the stress will be close to equilibrium stress which is considered to be the time-independent (plastic) contribution to the stress. Histories I - II provide an opportunity to study the influence of the stress level on the ratchetting behavior of the material.

AB - unloading to zero stress with the same strain rate magnitude.

B - mode switch to stress control. (Mode switch refers to the transfer of control from stress [strain] to strain [stress].)

BC - 1,000 cycles between zero stress and the ratchetting stress level conducted at a constant stress rate. In the tests the rise time $T_R$ to the ratchetting stress level was maintained constant. For each history one test for each of the rise times (2, 21 and 210 s) was conducted. Since the magnitude of the ratchetting stress varied slightly from test to test, the stress rate varied as well.
CD - loading to the ratcheting stress level with the same stress rate as used on BC.
DE - 700 s creep test introduced to see whether all the time dependence was exhausted after ratchetting.
EF - unloading to zero stress with the same stress rate as used on BC.
F - mode switch to strain control.
FG - loading to a strain of $\varepsilon_F + 1.00\%$ in strain control with the same strain rate as on OA. At point G the stress $\sigma_G$ was recorded.
GH - unloading to zero stress with the same strain rate as along OA.
H - mode switch to stress control.
HI - one cycle between zero stress and the stress $\sigma_G$ with a rise time of 2,100 sec. The purpose of this last test sequence is to see how the prior stress rate history and prior ratchetting stress level will affect the inelastic strain accumulation during this "slow" cycle.

The tests were performed at room temperature and at 550, 600 and 650°C. Due to the unexpected results of the elevated temperature tests not all rise times were performed at 550, 600 and 650°C. For the same reason the tests with History II were only conducted at room temperature.

**TEST RESULTS**

The results of the tests are summarized in Table 1. Stress-strain diagrams of individual tests are reproduced in Figures 1-4, with Figures 3 and 4 pertaining to 600°C.

An examination of these figures and Table 1 shows the following surprising results:

At room temperature the accumulated ratchet strain BC depends on the rise time $T_R$ for History I. As the rise time increases (the stress rate decreases) the accumulation of strain increases. Ratchetting is therefore influenced by the rate of loading. The accumulated strains are not negligible. The ratchet strain per cycle decreases from cycle to cycle. This can be best seen in Figure 2. Towards the end of the ratchetting program the strain advance per cycle is very small, a shakedown condition appears to have almost been reached. This trend is evident from Figure 5 where the inelastic strain (the strain at zero load) is plotted versus the number of cycles. This graph also demonstrates the rate (time) dependence of the inelastic strain accumulation. However, a subsequent creep test with the stress level equal to the maximum stress level of the ratchet cycle shows additional creep strain DE which decreases as the prior stress rate decreases. The capacity of the material to deform in a time-dependent fashion is not exhausted by the apparent shakedown during zero-to-tension loading. Upon a further tensile, strain controlled Pull (FGH), the material continues to harden as in a regular tensile test; the fact that the strain BF was accumulated during cyclic and creep loadings does not appear to matter. (The dashed extension emanating from A in Figures 1 and 2 is the postulated curve for an uninterrupted tensile test.) The subsequent slow cycle HI shows significant additional inelastic strain accumulation.

For History II and room temperature the accumulated strain for both the ratchetting and the creep periods are insignificant and can be neglected. This is evident from Table 1. A graph of ratchet strain versus cycles does not show any further accumulation. It appears that the relaxation drop has exhausted the capacity of the material to deform in a time-dependent (viscous) manner.
Most surprisingly the tests at 600°C do not show any significant ratchet or creep strains for either the 2 or the 210 s rise time (see Figures 3 and 4). Also the slow cycle HI does not show any significant accumulation of inelastic strain. It appears that type 304 stainless steel has lost its capacity for short-time, time-dependent deformation at 600°C. Since this result was unexpected, other tests were performed at 550 and 650°C. They again showed almost rate independent behavior (see Table 1). Because of these results no tests were performed with History II at elevated temperature. It is evident from Figures 3 and 4 that the stress strain curves have a ragged appearance, they exhibit serrated yielding. (The shape of the curves is not reproduced well since the sampling intervals for data recording were not small enough for such irregular behavior.) Their appearance is quite different at elevated and at room temperatures.

DISCUSSION

Normally one would expect that time dependence of deformation increases with increasing temperature. The fact that this was not observed in the present tests is attributable to the presence of strain aging which manifests itself by serrated yielding, inverse rate sensitivity or lack of it, and by an increase in flow stress with increasing temperature [14]. Serrated yielding and lack of rate sensitivity were observed in our tests and these results together with results reported in [15,16] suggest that strain aging is present in the elevated temperature ratchetting tests. It appears that this problem can be avoided by reducing the carbon content of the stainless steels. In recent years L-versions (low Carbon) of the stainless steels have been introduced.

Although the amount of inelastic deformation observed at room temperature may be surprising to many, this unusual behavior has been known for some time and was no surprise to the authors. In fact the conception of the test program was based on the presence of time (rate)-dependence observed in previous investigations [11-13]. These observations gave rise to the viscoplasticity theory based on overstress, VBO, which suggests that the stress is composed of time-independent (plastic) and rate-dependent (viscous) contributions (see [17,18]). By setting the stress level as was done in History I, the viscous contribution is nearly maximum and a rate-dependent ratchetting is observed (see Figure 5). If on the other hand the stress level is such that the viscous contribution to the stress is exhausted as in History II, the material should behave almost like a time-independent plastic material; this behavior was indeed observed, see Table 1.

The room temperature behavior was as expected and according to the concepts embedded within VBO. Accordingly it should be possible to predict the behavior using VBO. This task will be performed later. At the present time some encouragement exists since it was possible to predict the ratchetting behavior for a Ti-alloy which has very similar properties to those of 304 stainless steel at room temperature. Its ratchetting behavior was predicted with reasonable accuracy in [19].

The observations with Histories I and II also explain why the time-independent theories may underpredict the ratchetting behavior of stainless steels. These materials have, at least at room temperature, a significant rate dependence of their deformation. This part is not accounted for in the time-independent theories. Indeed if they would be applied to the conditions of ratchetting of History I, no ratchet strain and no creep strain would be predicted at all. The stress never exceeds the yield surface on the path AE in Figures 1 through 4. Because of this property, time-independent theory would be applicable to the computation of the elevated temperature results.
The present results are taken from a recent report [20] in which other facets such as the effect of prior cyclic hardening are discussed. The capability of the VBO in predicting cyclic hardening in uniaxial as well as in biaxial in-phase and out-of-phase loading is demonstrated in [19].

ACKNOWLEDGEMENT

The support of the Department of Energy made this study possible. Mr. E.J. Tracey of General Electric Company performed the heat treatment of specimens.

REFERENCES


12. D. Kujawski, V. Kallianpur and E. Krempl, "Uniaxial Creep, Cyclic Creep and Relaxation of AISI Type 304 Stainless Steel at Room Temperature," Journal of


---

**Figure 1** Stress Strain Diagram for a Rise Time of 2 s.

BC - Ratchetting strain, DE - creep strain.

Room temperature
Figure 2  Stress-strain Diagram for a Rise Time of 210 s.  
BC - Ratchetting strain,  
CE - creep strain. Room temperature.

Figure 3  Stress-strain diagram for a rise time of 2 s.  
BC - Ratchetting strain, DE - creep strain. Temperature - 600°C.

Figure 4  Stress-strain Diagram for a Rise Time of 210 s.  
BC - Ratchetting strain, CE - creep strain.  
Temperature - 600°C.

Figure 5  Inelastic Strain Accumulation During Ratchetting for Various Rise Times as a Function of Cycles. The effect of rise time is evident. Room temperature.
<table>
<thead>
<tr>
<th>Specimen Number</th>
<th>Test History</th>
<th>Temperature T (°C)</th>
<th>Rise Time $T_R$ (s)</th>
<th>Flow Stress $\sigma_A$ (MPa)</th>
<th>Stress Drop $\Delta \sigma$</th>
<th>Flow Stress $\sigma_R$ (MPa)</th>
<th>Stress Drop $\Delta \sigma$</th>
<th>Ratchetting Stress Level $\sigma_R^R$</th>
<th>Ratchetting Stress Rate $\dot{\sigma}_R$ (MPa/s)</th>
<th>Inelastic Strain Accumulated During Ratchetting BC, $\varepsilon_{RAT}$</th>
<th>Accumulated Creep Strain $\varepsilon_{CR}$</th>
<th>Combined Inelastic Strain $\varepsilon_{RAT} + \varepsilon_{CR}$</th>
<th>Flow Stress $\sigma_F$</th>
<th>$\sigma_G$</th>
<th>$\sigma_{G - \sigma_A}$</th>
<th>Inelastic Strain $\varepsilon_{SC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>I</td>
<td>20</td>
<td>2</td>
<td>217</td>
<td>-</td>
<td>217</td>
<td>108.48</td>
<td>0.561</td>
<td>0.365</td>
<td>0.926</td>
<td>1.810</td>
<td>257</td>
<td>40</td>
<td>0.702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>I</td>
<td>20</td>
<td>21</td>
<td>223</td>
<td>-</td>
<td>223</td>
<td>10.63</td>
<td>0.718</td>
<td>0.231</td>
<td>0.949</td>
<td>1.834</td>
<td>261</td>
<td>38</td>
<td>0.703</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>I</td>
<td>20</td>
<td>210</td>
<td>209</td>
<td>-</td>
<td>209</td>
<td>0.99</td>
<td>1.139</td>
<td>0.018</td>
<td>1.157</td>
<td>2.039</td>
<td>256</td>
<td>47</td>
<td>0.847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>II</td>
<td>20</td>
<td>2</td>
<td>235</td>
<td>40</td>
<td>195</td>
<td>97.56</td>
<td>0.001</td>
<td>0.004</td>
<td>0.005</td>
<td>0.877</td>
<td>260</td>
<td>25</td>
<td>0.653</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>II</td>
<td>20</td>
<td>21</td>
<td>213</td>
<td>39</td>
<td>174</td>
<td>8.29</td>
<td>0.008</td>
<td>0.005</td>
<td>0.013</td>
<td>0.899</td>
<td>238</td>
<td>25</td>
<td>0.731</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>II</td>
<td>20</td>
<td>210</td>
<td>285</td>
<td>73</td>
<td>212</td>
<td>1.01</td>
<td>0.002</td>
<td>0.004</td>
<td>0.006</td>
<td>0.876</td>
<td>306</td>
<td>21</td>
<td>0.653</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I</td>
<td>600</td>
<td>2</td>
<td>85</td>
<td>-</td>
<td>85</td>
<td>42.50</td>
<td>0.017</td>
<td>*</td>
<td>0.018</td>
<td>0.900</td>
<td>103</td>
<td>18</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I</td>
<td>600</td>
<td>210</td>
<td>84</td>
<td>-</td>
<td>84</td>
<td>0.42</td>
<td>0.020</td>
<td>*</td>
<td>0.026</td>
<td>0.949</td>
<td>104</td>
<td>20</td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I</td>
<td>650</td>
<td>2</td>
<td>92</td>
<td>-</td>
<td>92</td>
<td>46.00</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>0.886</td>
<td>117</td>
<td>25</td>
<td>0.060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>550</td>
<td>2</td>
<td>65</td>
<td>-</td>
<td>65</td>
<td>32.50</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>0.937</td>
<td>No Test Result in this Category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Inelastic strain accumulation is less than 0.01%.
LASER DIAGNOSTICS OF PACVD PROCESSES FOR DEPOSITING HARD FACE COATINGS

W. C. Roman, J. H. Stufflebeam, A. C. Eckbreth and F. A. Otter

United Technologies Research Center
East Hartford, CT. 06108 U.S.A.

ABSTRACT

Nonequilibrium rf plasma chemistry experiments using a longitudinal PACVD flow reactor are being conducted to improve the understanding of the mechanisms which govern the reactive plasma atomistic or molecular deposition of hard face coatings for erosion resistant applications. TiB$_2$ coatings were formed on Ti-6A1-4V substrates using TiCl$_4$ and B$_2$H$_6$ reactants. The reflective metallic coatings (≈1 μm thick) were smooth and the deposition rate was higher than generally reported levels. Analyses revealed the importance of substrate pretreatment/cleanliness and reactor contamination. Exploratory diagnostics include spectral emission surveys and narrowband, colinear, coherent anti-Stokes Raman spectroscopy (CARS) measurements. Since hardness and adhesion are critical parameters for characterizing tough, durable coatings, a state-of-the-art ultramicrohardness tester and custom built adhesion test apparatus are being implemented.

INTRODUCTION

Nonequilibrium reactive plasmas are recognized as a novel approach for a wide variety of material coating applications. This type plasma provides a unique environment wherein deposition of thin, hard face, conformal coatings can occur at much lower temperatures (critical for inhibiting stresses) and within more complex chemical environments than are possible by conventional vapor deposition techniques. A need exists for in-situ nonintrusive techniques for diagnosing and controlling PACVD systems used for thin coating deposition. If improvements can be made in durable coatings such as TiB$_2$, increased utilization of lightweight materials under advanced development will follow. A complete understanding of the process requires information on a large number of physical and chemical processes involving gas phase and gas-surface interactions and the associated synergistic effects. Knowledge of plasma species concentrations and temperature is required for correlation with the corresponding physical and chemical properties of the coatings. Table I is a partial list of compounds that have potential for tribological hard coatings. Included within the parentheses is the corresponding hardness in GPa. Denoted within the brackets is the thermal shock resistance as determined by the equation shown in the lower left. As a point of reference, cubic BN (hardest known compound) is shown with hexagonal TiB$_2$, which is ranked as the fourth hardest known compound. Detailed inspection and study reveal the titanium metalloids form an attractive group with additional advantages above those given in Table I.

Because of its superior erosion and thermal shock resistance, TiB$_2$ was selected as the initial hard face coating material for investigation. In addition,
the thermal expansion coefficient of TiB₂ (6.5 - 8.1 x 10⁻⁶/C range) closely matches that of Ti-6Al-4V (7.2 - 9.6 x 10⁻⁶/C range). Having a close match is favorable, but having a substrate with a slightly higher coefficient will result in some compression during cooling -- a desirable feature for the relatively brittle TiB₂. A TiB₂ coating of 58.88 wt.% Ti and 31.12 wt.% B is a very hard, simple hexagonal structure material, with a density of 4.52 g/cc and a melting point of approximately 3170 K. It is a very stable boride at room and elevated temperature, and the only boride of titanium that is stable in the presence of carbon; Ti-6Al-4V was selected as a substrate material. It represents a critical alloy of paramount interest in many commercial/aerospace applications. Its strength to weight and low temperature oxidation/corrosion resistance are excellent, but erosion and wear occurs rapidly, thus need for applying a durable hard face coating.

<table>
<thead>
<tr>
<th>CARBIDES</th>
<th>NITRIDES</th>
<th>OXIDES</th>
<th>BORIDES</th>
</tr>
</thead>
<tbody>
<tr>
<td>TiC (TiC)</td>
<td>TiN (TiN)</td>
<td>TiO₂ (TiO₂)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
</tr>
<tr>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
</tr>
<tr>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
</tr>
<tr>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
</tr>
<tr>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
<td>Ti₃B₄ (Ti₃B₄)</td>
</tr>
</tbody>
</table>

Table 1. Potential Compounds For Tribological Hard Coatings

APPARATUS

Figure 1 is a schematic of the PACVD reactor. The longitudinal, tube type geometry was selected as a design viable for later scale up to continuous production operation in addition to permitting high reactant gas flows and plasma power levels; increasing these parameters nominally results in enhanced deposition rates. The 4.0-cm-OD test section is constructed of high purity fused silica and high vacuum flanges for rapid disassembly. Up to 5 different reactant gases can be simultaneously injected at the inlet region. A 5 kW radio frequency (rf) power supply operating at 13.56 MHz is used to inductively couple the rf power into the plasma through a multiple-turn watercooled copper work coil. An rf shorting coil, located in the downstream region, is used for independent control of the plasma plume displacement relative to substrate position. Exploratory test results revealed the importance of minimizing contamination in the PACVD system as required for providing a high quality TiB₂ coating onto the Ti-6Al-4V substrate. As a result, a completely oil-free, venturi, vascorb, cryopump and vaceon pump system (see Figs. 2 and 6) is used including a MKS flow metering system with special traps (e.g., GaInAl) and getters to remove water vapor and oxygen. To augment pumpdown time and provide operating range flexibility, a 380 L/s turbomolecular pumping system, modified for corrosion resistant operation is used; this system is capable of initial operating pressures down to 10⁻⁷ torr. The cleanliness of the substrate prior to coating deposition is believed to be a key requirement for achieving good coating adherence to the substrate. Therefore, a test sample holder and isolated load-lock and transport system are interfaced with the PACVD system (see Fig. 1). This technique allows direct Secondary Ion Mass Spectrometry/Ion Scattering Spectrometry (SIMS/ISS) analysis of the substrates immediately following in-situ plasma cleaning and provides a unique capability not reported in the literature. To facilitate conducting
optical/laser diagnostic measurements from the reactive plasma zone to the tailflame region and adjacent to the substrate, optical ports and an automated traversing assembly are used. To conduct spectral emission survey measurements and to aid in identifying the strong spectral lines present, a 0.5m SPEX spectrometer system with an OMA was used, as shown in Fig. 1.

As hardness and adhesion are critical parameters for characterizing thin, hard face coatings such as TiB$_2$, a state-of-the-art ultramicrohardness tester and UTRC built adhesion test apparatus have been installed and preliminary calibration testing initiated. The ultramicrohardness tester (Nanoindenter) permits ultra-precision measurement of hardness, fracture toughness, adhesion and related properties of submicron volumes of material. With this instrument, the displacement of an ultra small diamond "indenter" is continuously monitored as shown in Fig. 3, as a controlled load is applied. The vertical position resolution obtainable is 0.2nm and the horizontal position resolution is 0.1µm. This will provide information on the uniformity of the surface hardness of the 13mm dia. substrates. A load resolution of 0.5µN permits indentations as small as 20nm to be produced repeatably. These figures reflect a level of resolution 10⁷-10⁸ times better than that of other mechanical microprobes. A shimadzu hardness tester (available within UTRC) would normally be

---

Figure 1. PACVD Reactor/Optical Emission Instrumentation

Figure 2. PACVD Overall System

Figure 3. Nanoindenter
used for evaluating these nominal µm thick TiB₂ coatings. However, several problems exist concerning interpretation of this type data particularly as it applies to relatively thin (<1 µm) hard TiB₂ coatings deposited onto softer substrates such as Ti-6Al-4V or pure Ti. Of prime concern is the fact that the "measured" hardness increases with decreasing indentation depth for very small indents. Another problem is interpretation of data from coatings such as TiB₂ that may exhibit a significant amount of elastic strain combined with elastic accommodation deformation by the substrate during deformation by the indenter. For these reasons, the nanoindentation is ideally suited for quantitative characterizations of the mechanical properties of the TiB₂ coatings having micron or submicron-size dimensions. The application of this technique will also be used for studying graded TiB₂ coatings and correlation of the coating hardness profile as a function of laser diagnostic measured PACVD parameters.

The adhesion or bonding characteristics of a thin hard face coating of the TiB₂ to Ti-6Al-4V are difficult to measure quantitatively. Among the possible techniques for measuring adhesion between a deposited coating and the substrate, one which has proven rather successful, is the scratch adhesion measurement. Using this technique, a diamond stylus (typically a Rockwell "C" indenting diamond) traverses the coating of interest under monotonically increasing applied force until coating/substrate decohesion occurs. An acoustic emission sensor is used to detect incipient coating failure. Monitoring the frictional force with a resistive-bridge-type transducer further enhances the detection of coating delamination and other failure modes. By using the combination of sensors, coatings less than one micrometer in thickness may be studied. UTRC studies will rely upon the use of a modified pin-on-disc apparatus, as shown in Fig. 4, with force transducer and an acoustic emission detector. The TiB₂ coated substrate will be rotated at the desired rotational speed (1-360 rpm range). A dc current driving an electromagnetic actuator applies precise normal forces throughout the 0-45N range. In series with the actuator and the specimen load pin is a highly accurate, multi-channel, force/moment transducer for measuring the applied normal force and resultant frictional (tangential) forces and/or moments. These signals along with the outputs from two optical rotary encoders, used to provide rotational position and velocity measurements, and the output from an eddy current probe serve as inputs to a Masscomp digital computer (5 KHz data rate). Full graphics analysis will be performed using the Masscomp computer in a post-process fashion. Currently, this apparatus is operational and in the final stages of calibration.
Figure 5 is a simplified schematic of the colinear phase-matched CARS system; the dye laser is scanned to generate the CARS spectrum. This scanned narrowband approach is used to get maximum species sensitivity by sacrificing temporal resolution. In this way, all of the available laser pump energy excites a single transition of the resonant molecule and increases the sensitivity by several orders of magnitude. The basic equipment, as shown in Fig. 6, includes a single-mode, frequency-doubled (532 nm) Nd:YAG primary beam pump laser and a narrow bandwidth (0.4 cm\(^{-1}\)) Stokes beam scanned dye laser (Lambda Physik Model 2002E). The two laser beams are aligned colinearly and focused in the medium. The frequency-shifted CARS signal is formed in the laser focus and all three beams exit the reactor where a dichroic separates the CARS beam from the incident wave mixing beams. Digital acquisition of the CARS and reference signal are obtained from fast PMT detectors. The digital acquisition system also records the wavelength shift of the scanned dye laser for calibration of the CARS spectrum. All signals are digitized and processed in a computer. As the dye laser is scanned, successive transitions of the molecular resonances are excited and modulate the intensity of the CARS signal. The result is

![](image)

**Figure 5. Colinear Phase-Matched CARS System**

Figure 6. Scanned Narrowband Colinear CARS System/PACVD System

a spectrum of the ro-vibrational manifold of the medium from which identification of species, their concentration and temperature (vibrational and rotational) can be
obtained. With appropriate referencing, the molecular concentration can be measured to within 1% of the total gas density and the population in the vibrational and rotational states determined for assignment of temperatures.

RESULTS TO DATE AND DISCUSSION

For the majority of experiments conducted to date the typical range of test parameters was: rf plasma power level, 0.2 - 2 kW; chamber pressure, 0.1 - 10 torr; total gas flow rates, 2 - 100 sccm; and test times, 5 - 30 min. Detailed analysis of the initially deposited TiB$_2$ coatings using Auger/SIMS/ISS/ESCA revealed the presence of submicron carbonaceous inclusions and measured impurities of C, O and N. On the surface, the oxygenated Ti surface compounds were measured to be predominantly TiO$_2$. The requirement for starting with contamination free substrate surfaces is recognized as important for good adherence of coatings. The original standard metallurgical procedures used for preparing both the Ti-6Al-4V and reference high purity Ti substrate samples were determined to be unacceptable. An added step of electropolishing with 5% perchloric acid in methanol was used and proved to be effective. ISS-SIMS data taken at each step of treatment revealed that significant amounts of aluminum left on the samples from polishing was removed after the electropolishing. ISS-SIMS results on plasma treated samples showed that a hydrogen plasma is effective for removing elements that form volatiles, e.g., B, O, C, etc. and that the degree of cleaning can depend on sample position in the PACVD reactor relative to the plasma discharge. In general, locations near the plasma plume tail gave the best cleaning. Figure 7 is an example illustrating the effectiveness of the plasma cleaning step. Neon SIMS measurements are shown for both the before plasma cleaning (solid curves) and the after plasma cleaning (dashed curves) steps. Corresponding test conditions were: 45 sccm argon, 25 sccm hydrogen, 20 torr, 0.5 kV and a test time of 30 minutes. The hydrogen plasma significantly reduced the B, Na, and Al. Future experiments using substrate bias may be necessary to sputter clean any residual Na or K. The availability of high purity reference samples of Ti-6Al-4V, Ti and TiO$_2$ (rutile) has also permitted more detailed investigation of the impurity concentration profiles.

Figure 7. Neon SIMS Data of Plasma Substrate Pretreatment

Approximately 30 candidate species, including compounds/radicals, that may be present in the plasma or downstream region and associated with the TiCl$_4$ and B$_2$H$_6$ reactants have been documented and information collected on the associated spectroscopic data where available. These species include B, Ti, Cl, H, BH, BH$_2$, BCl, TiH, BH$_2$, Ti$_2$, B$_2$H$_6$, TiCl, HBCl$_2$, TiB, B$_2$, TiB$_2$, Ti-B, Ti$_2$B, Ti$_4$B$_4$, B$_2$H$_6$, B$_5$H$_9$, B$_4$Cl$_4$, BCl$_3$, BCl$_2$, BCl, H$_2$, HCl, H$_2$BCl, TiCl$_4$, TiCl$_3$, and TiCl$_2$.

Figure 8 is an example of the BH molecular band emission measured in the 425-440

---

82
nm range when operating a plasma on only diborane (B\textsubscript{2}H\textsubscript{6}) at 6 torr and 0.2 kW. The Q transitions at 433.16 nm can be noted; it is from this type of data that the rotational temperatures of the plasma can be calculated.

![Figure 8. PACVD Plasma BH Emission](image)

Figure 8. PACVD Plasma BH Emission

Figure 9 is another example of the plasma emission measured with 0.1 nm resolution when operating a B\textsubscript{2}H\textsubscript{6}/TiCl\textsubscript{4}/Ar plasma at 0.2 torr and 0.25 kW. The strong atomic lines have been identified with the majority being Titanium I or ArI. At the present time it is not known how many of the total possible candidate compounds, including neutral or ionic species, even exist in the gas phase, much less have a characteristic, measurable spectrum. As an example, Figure 10 illustrates molecular band structure in the 346-446 nm range that has been identified as TiCl. All lines in this figure have been identified. In this case, the plasma was operating on Ar and TiCl\textsubscript{4} at 0.45 torr and 0.45 kW. Work is continuing on identification of the atomic lines and band structure over a wide wavelength range as observed from the PACVD plasma.

![Figure 9. PACVD Plasma Emission](image)

Figure 9. PACVD Plasma Emission
Emission and absorption spectroscopy need accessible optical transitions, therefore for these type PACVD plasmas a more general technique applicable to all molecules is required. All molecules have at least one Raman active vibrational mode, therefore Spontaneous Raman Spectroscopy (SRS) is a candidate technique. It is a fairly simple technique and spatially precise; unfortunately, it is an extremely weak process. It also suffers in situations with high background luminosity or in trying to achieve good sensitivity at relatively low pressures. To overcome these disadvantages, the colinear narrowband scanned CARS technique is being applied to tests in progress.

From the experimental results obtained to date in this continuing research program, it appears that no one diagnostic technique will suffice, but only through the complementary use of different techniques will a fundamental understanding of the relationship between nonequilibrium plasma process variables and properties of the deposited coatings evolve. From this type basic knowledge, plasma chemistry induced mechanisms can be formulated and subsequently a much needed predictive capability developed. Research focused on these aspects is continuing.

Acknowledgement

Part of this research is sponsored by a DOE grant (No.86ER13560) that addresses applying novel laser diagnostic techniques to PACVD materials processing. The authors gratefully acknowledge the contribution of their colleagues at UTRC.

References


DEVELOPING DIAGNOSTIC METHODS FOR PLASMA-PARTICLE INTERACTION

C. B. Shaw, Jr.
Idaho National Engineering Laboratory/ EG&G Idaho, Inc.
Idaho Falls, Idaho 83415-1625, U. S. A.

ABSTRACT

Three diagnostic methods for study of plasma-particle interactions are described. The first, the single-line method of emission spectroscopy for measurement of plasma temperature, is well established but required careful implementation of automatic methods for data reduction and analysis, and will need to be extended at a later date. The second, laser-induced fluorescence (LIF) for measurement of density of species vaporized from the particles, has been found to require substantial modification in experimental detail and data analysis from its origin in flame diagnostics. The third, Doppler-shifted Rayleigh scattering for direct measurement of plasma velocity, is a new undertaking suggested by a single published report of successful measurements on a very high current arc.

INTRODUCTION

This is a report on work in progress -- not brought to a definitive conclusion -- at the Idaho National Engineering Laboratory (INEL). It is part of the investigation of interactions between a plasma and particles entrained in the plasma which is being conducted jointly by the INEL and the Massachusetts Institute of Technology (MIT). The project addresses the national need for technological innovation which can radically reduce the costs to American industry of, for example, extracting metal from its ore and converting it to useful form. The long range goal is a paradigm for the rational design of thermal plasma generators and processors to give optimum performance in any specific application. Such a paradigm requires creation and exercise of computational models of the plasma processing operation, generation of a body of reliable experimental data from diagnostic measurements of the plasma processing system, and critical comparison and interpretation of predictions and measurements to validate the models.

The INEL is pursuing a quantitative description of the heat, mass, and momentum transfer associated with particles immersed in a thermal plasma. Methods are being developed to measure plasma properties adjacent to the particles. Experimental results are compared with predictions of MIT's computer models. Interpretation of these comparisons is the product of close collaboration between the INEL and MIT investigators in which results obtained at each institution influence the course of research at the other.

The three diagnostic methods reported here are in three distinct stages of development. The method used for temperature measurement, single-line emission spectroscopy, is old and generally regarded as well established, even though
limitations are pointed out which will require its modification or replacement later. The method under development for vapor density measurement, laser-induced fluorescence (LIF) is well established in combustion research, but had never been used at plasma temperatures before. The higher collision frequency in plasmas has been discovered to introduce a phenomenon not reported in flames, which requires more sophisticated data analysis methods. Finally, the method proposed to measure plasma velocity is barely beyond the concept stage, with a proof of principal experiment just being implemented.

TEMPERATURE MEASUREMENT BY SINGLE-LINE EMISSION SPECTROSCOPY

The single-line method is well known and widely used because of its apparent simplicity [1]. A table look-up is performed to find (for a given pressure) the temperature, T, at which a chosen spectral line exhibits an observed absolute emissivity, e. (Similar use may be made of continuum radiation in a selected wavelength interval.) Since each spectral line has a characteristic temperature, T*, at which emission is a maximum, T is a double-valued function of e and one must know (generally by assumed monotonicity) whether the temperature is above or below T*. Representative data, obtained at the INEL by the conventional single-line method, are shown in Figure 1 as a function of distance from the orifice along the plasma torch's axis of symmetry. The close comparison with predictions of the MIT (Szekely and Dilawari) model is evident. Figure 2 makes similar comparisons for temperature profiles at three axial positions.

As a test of the precision of this method in the present application, similar measurements were made using spectral lines of neutral and ionized argon, and also the argon continuum. By the standards of spectroscopic plasma temperature measurements, the agreement shown in Figure 3 is quite good. One expects the Ar\textsubscript{I} measurements to be more accurate near the peak temperature observed, because the emissivity of the ionized argon line is still increasing rapidly with temperature. By contrast, as shown in Figure 4, both the neutral argon line and the continuum are approaching their respective T*\textsubscript{S}, where the curve is flat, so that a small error in emissivity produces a large temperature error. On the other hand, at lower temperatures, measurements using the argon ion line suffer signal to noise problems.

The single-line method necessarily assumes that the composition of the plasma is known at the instant the experimental data are recorded, in order that the functional relationship between e and T be known from either theory or prior experiments. Aside from the dependence of the emissivity on the partition function for the radiating species, which is altered by the presence of impurities, e is proportional to the total density (in all energy levels) of the particular species, e.g. neutral argon or singly-ionized argon. The emissivity therefore changes (noticeably, in the case of a weakly-ionized plasma, for the ionized species) with the degree of ionization of the plasma. The latter is strongly influenced (according to the Saha equation [1]) by rather small concentrations of elements with appreciably lower ionization potential than the radiating element, e.g. metal vapor in an inert gas. Radiation from the more easily ionized element has, however, been used to measure temperature by the single-line method in an impure plasma [2].

A more robust temperature measurement technique may be purchased at the cost of experimental complexity. The multi-line emission spectroscopy method is based on comparison of the relative strengths of several spectral lines [1]. When all lines are from the same species, the dependencies on both partition function and species density cancel out of the temperature determination. The lines can be measured either sequentially [3,4], if the plasma is sufficiently stable, or simultaneously [5].

When it becomes desirable to measure temperature in a plasma which is heavily loaded with particles, a choice will be made between the single-line method with
Figure 1. Comparison of MIT model predictions (solid curve) and INEL measurements (circles) of temperature versus axial distance along centerline for 700 A current.

Figure 2. Predicted (MIT) and measured (INEL) temperature versus radius at three axial positions for 700 A current.

Figure 3. Temperature measured by Ar line at 430.0 nm, Ar line at 480.6 nm, and by Ar continuum at 430.8 nm, all versus radius at axial position 1 mm from orifice for 500 A current.

Figure 4. Emissivity versus temperature for Ar line at 430.0 nm, Ar line at 480.6 nm, and Ar continuum at 430.8.
metal vapor as the measured radiation source, or a multi-line method. The latter is more nearly immune to error due to scattering of the radiation from particles, since the wavelength dependence of the scattering is weak, and the absolute intensity at each wavelength is not important.

Each of the above spectroscopic methods assumes knowledge of the local emissivity, but that is not the quantity which is directly observed. Instead, one records the integrated intensity, $I$, emitted in a certain direction by all sources within an infinite cylinder, as indicated in Figure 5. Whether one records the data with monochromator and photomultiplier, for each transverse position, $x$, [3], with monochromator and optical multichannel analyzer (OMA) for an entire cross section as is now being done, or with interference filters and imaging devices to record entire images of the plasma [5-7], one is faced with an example of the problem of reconstruction from projections. In the general case the recorded intensity is a function of the orientation angle, $u$, as well as $x$. (It is also a function of $z$, the direction orthogonal to $x$ and $y$, but that is suppressed since the problem can be solved independently for each value of $z$.) The equation which must be solved for the emissivity $\varepsilon(r,\theta)$ is

$$I(x,u) = \int \varepsilon(r,\theta) \, dy$$

where

$$x = r \cos(\theta+u), \quad y = r \sin(\theta+u).$$

In the special case that $\varepsilon$ is independent of $\theta$, so that $I$ is independent of $u$, Eq. (1) reduces to the more familiar Abel integral equation. The general case, Eq. (1), can be solved when $\varepsilon(r,\theta)$ does not differ too radically from a Gaussian or exponential function of $r$ [8]. However, both this general problem and the ordinary Abel inversion are "ill-posed" problems by Hadamard's definition (see, for specific application to the Abel problem, Reference 1). Numerical errors accumulate unless great care is taken, for example, by use of the method of best accessible estimation [9].

At the present time, with only light particle loading of the plasma, or no loading at all, the assumption that the plasma is axially symmetric is sufficiently accurate. The analytical method currently in use is not original to this task and has been described elsewhere [10].

**METAL VAPOR DENSITY MEASUREMENT BY LIF**

Laser-induced fluorescence uses a laser tuned to resonance with a certain spectral transition in the species whose density is to be measured. The laser is focused to minimum volume and the power increased until saturation is achieved -- according to the principle of detailed balance, equal populations in the upper and lower energy levels of the transition. (Saturation is not essential, but makes the data independent of fluctuations in laser power, and simplifies data analysis). Inelastic collisions of the excited atoms populate all members of a multiplet containing the upper energy level which was pumped, and measurement of the absolute intensity of allowed radiative transitions from the various multiplet levels to members of a lower multiplet make it possible to calculate the initial density in the specific lower energy level, since the fluorescent intensity is proportional to that density. The analytic details developed by Winefordner and others for flame diagnostics are reviewed in [11] for the case of an atom with an arbitrary number of energy levels. That paper reported density distributions of neutral and ionized manganese, vaporized into a welding plasma from a stainless steel anode, which peaked at a surprising distance from the axis of the plasma. The puzzle was resolved when Snyder discovered that, in a plasma, fluorescence is induced by the laser pulse not
only from the intended multiplet but from many (if not all) others as well [12].
This is currently assumed to be due to the equilibrating effect of the much higher
collision rates, especially with electrons, which atoms experience when in a plasma
rather than a flame. The question then arises, is the plasma in the laser focal
region in partial local thermodynamic equilibrium, with only the multiplets involved
in the resonant transition not in equilibrium? If so, not only the metal vapor
density, but its temperature, can be determined from the LIF data after sufficient
analysis.
The experimental system has been slightly modified to record as much of the fluorescent spectrum as possible simultaneously, to make minimum demands on plasma stability. Boltzmann plots of the data are being made to test the partial equilibrium possibility. If such a great simplification is not possible, then one must deal with the entire coupled set of transition rate equations on some level of approximation. If possible, the time-dependent problem should be integrated numerically, to allow deconvolution of the laser pulse length from the data, since the (ten nanosecond) pulse length of the laser now available is comparable to the decay time. A flash-lamp pumped laser, with millisecond pulse length, would simplify this analysis somewhat.

PLASMA VELOCITY MEASUREMENT BY DOPPLER-SHIFTED RAYLEIGH SCATTERING

A literature search reveals no fully-documented method for accurate measurement of the flow velocity field of a plasma with the temperature, density, and maximum flow rates expected in an industrial processing plasma. However, there has been a report that plausible flow velocities were measured in a very high current (900 kA) arc by means of the Doppler shift (proportional to velocity component along the observation direction) in laser light Rayleigh scattered from the atoms and/or ions of the arc [13]. To improve resolution of the minute shift in wavelength expected, stringent methods were used to narrow the bandwidth of the laser, but the authors still found it necessary to deconvolve the response function of their spectral measuring instrument, a scanning Fabry-Perot interferometer, from the data before the thermal velocity distribution, with the small Doppler shift superimposed, could be discerned. Since deconvolution, like Abel inversion, is an ill-posed problem, the precise deconvolution algorithm used is an important issue. Unfortunately, this critical detail was referenced by a paper "to be published" which has never appeared. Therefore, it is planned to use a deconvolution method which has been demonstrated successfully both on simulated data (complete with simulated background and instrumental noise) and on experimental data [14].

The principal background noise to be overcome has been estimated to be continuum radiation from the plasma. To avoid a ruinous spectral line background, the initial measurements will be made on a nitrogen plasma, rather than an argon plasma, since it is an argon ion laser that is available. This is a cw laser of a few watts output, so the estimated signal to noise ratio is about -10dB, well within the ability of available instrumentation (the OMA and/or a beam-chopper lock-in amplifier combination) to dig the signal out of the noise. The laser has an internal etalon for longitudinal mode selection, which aids in bandwidth reduction, the consideration which ruled out an available pulsed laser.

After careful consideration of commercially-built Fabry-Perot interferometers, and consultation with the authors of recent publications reporting their use for high-precision interferometry, it was decided that no single unit would quite give us the finesse (ratio of bandwidth over which measurement can be made without ambiguity -- the spectral free range -- to resolution) needed to make resolution of the Doppler shift possible while still covering the entire line shape. A multipass scanning tandem Fabry-Perot was therefore ordered built to our specifications. The necessary electronics, vibration-isolation and thermal control devices have been ordered, and will be in place by the scheduled delivery date of the instrument. The laser spectrum, convolved with the instrumental response function, will be measured using unscattered laser light. That shape will then be deconvolved from the measured scattered light to reveal the Doppler shift due to combined thermal and flow velocities. Resolution is estimated to be adequate to reveal a flow velocity as low as 100 m/s. Indirect estimates of the flow velocity from our torches suggest 500 m/s, which should be sufficient for proof of principal.
ACKNOWLEDGMENTS

The author is grateful to his collaborators at the INEL, James A. Batdorf, Brent A. Detering, Patricia Hembree, Larry D. Reynolds, and Stuart C. Snyder, and to Prof. J. Szekely and A. H. Dilawari at MIT. This work was supported in part by the United States Department of Energy, Office of Basic Energy Sciences, Division of Engineering and Geosciences, under Contract Number DE-AC07-76ID01570, and in part by the United States Navy, Office of Naval Research, Code 471, under government order No. N00014-84-F-0152.

REFERENCES


HIGH TEMPERATURE PARTICLE PROCESSING

J. F. Elliott, R. E. Spjut, and P. Bolsaitis

Department of Materials Science and Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139, U.S.A.

ABSTRACT

An apparatus has been designed, built, and placed in operation that permits study of gas-particle reactions and physical and chemical changes that take place when a particle is heated and cooled rapidly. The time resolution for an experiment is approximately 1 millisecond, and the possible temperature range for the experiments is between 1,100 and 4,000 degrees K. The particle is supported in the reactor by the interaction of an electrical charge on the particle and a vertically oriented DC potential. The particle is heated with a 20 w CO\textsubscript{2} laser, and the system is designed to process the particle for conditions simulating those encountered by a particle in passing through the flame of a thermal plasma. Experiments are in progress on phase transformations that may occur on heating and cooling of alumina, silica, carbonized alumina, iron-nickel alloys, and the ignition and burning of various types of carbon. The program includes work on problems encountered in measuring the temperatures of small particles that are being heated and cooled rapidly at rates of up to 1 million degrees per second, and where there are phase changes in a material.

INTRODUCTION

A novel apparatus for study of gas-particle reactions at high temperatures, and the chemical and physical behavior of small particles at high temperatures and during rapid changes in temperature, has been built and placed in operation. The particles to be studied can be of virtually any material and in the size range of between 10 and 150 micrometers. The apparatus is based on the famous Millikan oil drop experiment in which a particle carrying an electrical charge is suspended by means of an electrostatic field. While suspended, the particle can be heated by a split laser beam, and particle temperatures of up to 4,000 K have been reached. The reactor is sealed, and the composition of the atmosphere surrounding the particle can be controlled; however, easily ionized gases must be excluded. Materials that have been studied in the apparatus are alumina, silica, carbon-coated alumina, iron-nickel alloys, and various types of carbon.

THE APPARATUS

The particle is suspended between two vertically oriented electrodes across which a DC potential is applied. A ring electrode carries an AC voltage which restrains the
lateral movement of the particle. The arrangement of the electrodes is shown in
Figure 1. The electrodes are black anodized aluminum, and they are housed in a delrin
shell. The electrodes and shell comprise the reactor. The chamber contains a number
of ports; two of them are opposed on a horizontal diameter for admission of the split
laser beam for heating the particle. There are also gas inlet and outlet ports, and
ports for the position control system, for temperature measurement, for observation
of the particle with a microscope or imaging it with a high-speed camera, for introduc-
tion of the particle, and one on the vertical axis of the lower electrode for captur-
ing the particle after an experiment.

The particle is heated by a split beam from a 20 w CO₂ laser. The two beams are
focused on the particle from opposite sides, and the intensities of them must be
adjusted so that the thermophoretic effects from opposite sides of the particle
cancel. The temperature of the particle is sensed by three photomultiplier tubes
(550, 750 and 850 nanometer radiation) by means of a bundle of optical fibers that
convey radiation from the particle to the tubes. The vertical position of the
particle is determined by an optical system which consists of a set of lenses, an
image intensifier, a double fiber optic cable, two photomultiplier tubes, and an
electronic circuit to sense the output of the photomultiplier tubes. The system
senses whether the particle is above or below the desired position, and through a
software package in the dedicated computer (PDP-11/73), the DC voltage is adjusted to
maintain the particle in the desired position. The particle is illuminated with a
HeNe laser (0.002 w) to aid in injection of the particle into the chamber, and for
position control. Through software in the computer, the output of each of the
temperature photomultiplier tubes is recorded at the rate of 5 kilobytes per second,
and the output of the CO₂ laser is controlled as desired. The response times of the
temperature sensing and position control systems are each less than 0.1 milliseconds.
Figure 2 is a schematic diagram of the layout of the system on its optical bench.

An experiment is started by injecting a few particles into the reactor with an
hypodermic syringe. The particles are charged by friction in passing through the
needle. By manipulation of the AC and DC voltages on the electrodes, one of the
particles is selected for the experiment and the others are excluded from the reaction
field. The CO₂ laser is then pulsed to heat the particle by a command to the
computer, and the computer is programmed to control the output of the laser.

Work is in progress to develop means for capturing a reacted particle for
examination which will be principally by electron imaging instruments. One means of
capture is to suck the suspended particle into a quill inserted into the reactor. The
particle is directed to lodge on an SEM grid mounted at the end of the quill. The
second is to aspirate the particle into the small hole drilled on the axis of the
lower electrode. An SEM grid is positioned in the opening for capture of the
particle. It is to be noted that capturing one particle, and examining that particle
is a much more taxing exercise than the general problem of sampling a cluster of
particles and examining the group one by one.

MATERIALS STUDIED

Time-temperature profiles (i.e., thermograms) for conditions of rapid heating and
cooling have been obtained for a variety of materials as is discussed below. These
measurements have served two purposes: First, they provide information on the phase
changes and oxidation (or combustion) rates of small particles of various materials;
and Second, they have permitted us to evaluate the capabilities of the thermobalance
for a variety of experimental conditions. Experiments that have been conducted on
materials are discussed briefly below.

Alumina and Carbon-Coated Alumina - Pulsed heating experiments have been con-
ducted on 50 micrometer alumina particles which have been raised to temperatures
Figure 1 Exploded View of Chamber Electrodes in Electrodynamic Balance.

Figure 2 Apparatus on Optical Bench.

Figure 3 Apparent Temperature of Pulse Heated Al₂O₃ Particle, 50 µm dia., recalescence at 21 ms.
up to 3,500 K with melting and subsequent solidification of the particle, the latter taking place over an interval of less than 0.2 millisecond. A typical time temperature profile that was obtained is shown in Figure 3. Similar experiments have been conducted on carbon-coated alumina particles in an oxidizing atmosphere. The heating rates are considerably enhanced over those obtained with uncoated alumina, partly because of the greater absorbivity of the particle and partly because of the heat from oxidation of the carbon. This work on alumina-carbon composites is continuing both in oxidizing and reducing atmospheres to study the carbon-alumina reactions at high temperatures.

Silica - Particles of 50 to 100 microns in size have been heated to temperatures above 2,500 K. The allotropic transformations and melting in silica have been recorded on the thermograms. The pattern of emitted radiation as a function of time, as illustrated in Figure 4, results from the complex interrelationship between the several phase transformations and the different optical properties of the phases involved. This work is continuing with the aim to understand better the mechanisms and rates of phase transformations, and decomposition of silica in reducing environments.

Fe-Ni Alloys - The kinetics of solidification of iron-nickel alloys in inert and oxidizing atmospheres is being studied in a separately funded program. Of particular interest is the effect during the solidification process of minute particles within the metal of iron oxide, alumina, etc., on the nucleation of the solid metal phase. One aim of the work is to eliminate the residual oxide within the metal particle when it is liquid by in situ reduction. The methods of studying the microstructure of these particles is in progress, as noted above. Figure 5 illustrates a typical set of melting-solidification curves obtained. The three traces are for outputs of the photomultiplier tubes that have been normalized at the melting point of the alloy. The amount of recalescence (at 147 milliseconds) observed is an indication of the relative effectiveness of oxide particles within the particle or on the surface of the particle as nucleation centers for the formation of a crystalline metallic phase. It has been observed that the recalescence when the atmosphere in the reactor is nitrogen is approximately 250 degrees, but it is only approximately 150 degrees when the particle is exposed to air. It is assumed that the lower value results from an oxide forming on the surface as a result of reaction of iron with oxygen in the air. The oxide can serve as a heterogeneous catalyst for nucleation of the solid phase of the alloy.

Oxidation (Combustion) Reactions - Study of the kinetics of oxidation reactions at high temperatures in plasma-like environments has been begun with work on the oxidation of tungsten and various types of carbon. The latter is supported by a separate program. The oxidation reactions have a number of similarities to combustion reactions. What is known about the combustion of carbon, and the results of our experiments on the kinetics of combustion reactions enhances our ability to conduct and interpret reactions by which metal and alloy particles are oxidized or nitrided in plasma systems. Figure 6 is a typical time-temperature profile for a combustion reaction. The laser beam ignites the particle and then the temperature of the particle is influenced strongly by the heat released from the oxidation reaction. In these experiments which have been carried out at temperatures below 2,000 K, extended reaction times can be attained, and heat and mass transfer models have been developed and tested against the experimental data for the combustion in air of spherocarb particles. In Figure 6, the particle was heated briefly by the laser beam and ignition occurred at 40 milliseconds. At that time the laser beam was turned off. Subsequently the temperature of the particle rose, and then it apparently declined. The significance of the drop in the curves after approximately 70 milliseconds is discussed in the next section.
Figure 4  Pulse Heating of 100 μm SiO₂ Particle, 550, 750, and 850 nm Color Temperatures. "1", "2", and "3" tentatively are phase transitions: (1) α-quartz --> β-quartz, (2) β-quartz --> tridymite, (3) tridymite --> cristobalite.

Figure 5  Three PMT Thermograms for a 45 μm Fe-30% Ni Particle in Nitrogen. Recalescence at 148 ms.
Figure 6 Single-Color Temperatures Calculated from Intensity Measurements Made on a Burning Laser-Heated Char Particle.

Figure 7 Thermograms for an Alumina Particle. The outputs of three photomultiplier tubes (550, 750, and 850 μm) have been converted to temperatures normalized at melting point of alumina.

Figure 8 Experimental Results for a Temperature Change Measured by a Thermocouple, Two Single-Color Pyrometers, and a Two-Color Pyrometer. Left, temperature increase. Right, temperature decrease.
EMISSIVITIES AND TEMPERATURE MEASUREMENTS

Two areas of primary interest for the project are the techniques and methods for temperature measurement of single particles, particularly in the range of 2,000-4,000 K range. The emissivities of many materials are not well known at temperatures above 2,000 K. The problem is of particular importance with materials that are partially absorbing to the impinging laser heating beam, alumina being of interest in the studies now in progress. Figure 7 shows three simultaneously measured thermograms of a 50 micrometer alumina particle. The laser beam was turned on at 130 milliseconds, and it appears that the particle moved out of the beam at approximately 137 milliseconds. There is one thermogram for each photomultiplier tube, and each has been normalized at the melting point of alumina (2323 K), and the emissivities of liquid alumina in the three wavelengths were assumed to be the same. The maximum temperatures calculated from the outputs of the three photomultiplier tubes differ by several hundred degrees, but it is to be noted that the apparent disagreement is very much smaller at the recalescence peak. Then there is a large disparity between the 850 nm trace and those for the other two wavelengths at lower temperatures. It is concluded from the results of a number of measurements, and from data available in the literature, that the emissivities of both solid and liquid alumina vary appreciably with wavelength of the radiation, and that for each wavelength the emissivities may also vary with temperature. As a result of this, two-color temperatures are also not reliable unless the emissivities for the two wavelengths are known. The temperatures obtained with the 550 nm photomultiplier tube appear to be the most reliable.

The response of single and multiple color pyrometric detectors for rapidly heated and cooled particles has been analyzed to optimize methods of temperature measurement and control. Figure 8 shows the response of single and two color pyrometers to the temperature rise and fall observed in the pulse heating of a small Pt-Rh thermocouple. The two-color response follows the true rise in temperature of the thermocouple bead very well, and it is superior to the results of either single color trace. It is also much superior to a first order response curve. With a decrease in temperature of the bead, the single color responses are better than that of the two-color response, but none is as good as a first order response. The lag in the recorded curves on cooling is very much larger than that encountered with particles in the thermal balance (See Fig. 3) because of the very much larger size of the thermocouple bead. The characteristics of this response are, of course, critical in the calculation of temperatures from radiation intensities emitted from small particles both in the thermal balance and in other high temperature systems. Understanding of these phenomena is particularly important at temperatures above 2,000 K where no independent temperature calibration can be established and the optical properties of materials are not well known.

The behavior of the curves in Figure 6 after 60 milliseconds illustrates another issue to be considered in determining the temperature of a particle. The apparent decline in the temperature of the particle arises because the particle is decreasing in size, and the total radiation emitted in each of the wave lengths decreases as well. The particle is pure carbon so that its emissivity at all wave lengths is nearly one. "Two-color" temperatures calculated from each of the three pairs of radiation curves are in good agreement, and they show that the temperature of the particle actually remains almost constant until approximately 100 milliseconds at which time the temperature drops rapidly because the rate of loss of heat from the small particle by convection overwhelms the rate at which heat is supplied by combustion.
This research program is directed towards understanding the role of quasi deterministic large scale vortices in the dispersion of particles by free turbulent shear layers. The primary objective of this experimental study is to determine the influence and importance of these vortex structures on the particle dispersion process over a range of particle and flow parameters. It has been hypothesized that if the time scale associated with the vortex motion and the particle relaxation time are of the same order, significantly enhanced particle dispersion may occur.

NOMENCLATURE

\( u \) = downstream velocity fluctuation, rms value
\( U \) = downstream mean velocity
\( U_H \) = velocity of high speed stream
\( U_L \) = velocity of low speed stream
\( x \) = downstream distance from trailing edge of splitter plate
\( y \) = cross stream distance from splitter plate
\( \tau_A \) = solid particle aerodynamic response time
\( \tau_f \) = time scale of large vortex structures
\( \lambda \) = velocity ratio parameter, \((U_H - U_L)/(U_H + U_L)\)
\( \delta \) = vorticity thickness of mixing layer

INTRODUCTION

This research program is concerned with evaluation of the importance of organized vortices in the dispersion of particles in free shear flows. Turbulent flows of this type have been studied extensively (Ho and Huerre 1984). The primary objective of this experimental study is to determine the influence and importance of quasi-deterministic vortex motions on the particle dispersion process over a range of particle and flow parameters. Crowe and Troutt (1983) and Crowe, Gore and Troutt (1985) have suggested that if the time scale associated with the organized vortex motion and the particle relaxation time are of the same order, significantly enhanced
particle dispersion may occur. Past experiments (Yule, 1980) and recent numerical simulations (Crowe, Gore and Troutt, 1985, Gore, Crowe, Troutt, and Riley 1985, and Chung and Troutt 1987) lend strong support to this idea.

The experimental program employs high speed photographic methods to track the dispersing particles. The nature of the instantaneous flow field is measured using an array of hot-wire sensors. By measuring the particle dispersion and the large scale flow features direct cause and effect relations between the vortex motions and the particle dispersion process can be ascertained. In addition laser velocimetry techniques will be used to further check and support both the photographic particle dispersion results and the hot-wire flow measurements.

A specific objective of this research is to explicitly determine relationships between the organized vortex patterns in the mixing layer and the particle dispersion process. One parameter of interest is the ratio of time scales between the vortices and the particles as previously mentioned. Other important phenomena which will also be experimentally evaluated are the effects of vortex pairing and the effects of three-dimensional motions in the mixing layer on the particle dispersion. The initial measurement program is concerned primarily with low particle concentrations such that the effect of the particles on the flow field will be small (one-way coupling). Later efforts will investigate the more complicated problem associated with high particle concentrations, where two-way coupling is operative.

The anticipated result of this study is a new physical model for particle dispersion in turbulent shear flows. The results will also be useful in directing the application of current two- and three-dimensional time dependent turbulence simulation models (A. Leonard, 1985, Hussaini and Zang, 1987) to the particle dispersion problem. These new computational models are needed to accurately predict particle dispersion in shear flow turbulence. Initial efforts along this line (Gore, Crowe, Troutt and Riley, 1985 and Chung and Troutt, 1987) have already shown significant potential.

**EXPERIMENTAL APPARATUS AND TECHNIQUE**

An open return wind tunnel with a test section of cross-section 45.7 cm x 60.96 cm and of length 182.9 cm is used for the experiments. A schematic of the tunnel is shown in Figure 1. The tunnel is fitted with a splitter plate which enables the independent control of the free stream velocities on either side of the plate. The free stream velocities are altered through the insertion of flow resistance screens upstream of the flow management section. Free stream velocities from 2 to 25 m/sec with high quality flow conditions are producible with this facility. Typical turbulence intensity levels at a free stream velocity of 10 m/sec are approximately 0.1% or less with mean flow uniformity within ±1%.

By controlling the free stream velocities on either side of the splitter plate a range of the velocity ratio parameter, $\lambda$, can be maintained. Both plane wakes, $\lambda = 0$, and mixing layers, $\lambda > 0$ can thus be investigated with this facility.

Solid glass bead particles can be introduced into the shear flow using a recently constructed particle injection device shown in Figure 2. This device uses a combination gravity feed with air transport concept to achieve adjustable particle flow rates with both continuous and intermittent injection modes. Particles with diameters from 10 μm to 200 μm can be injected in a controlled fashion using this device.

Presently, a smoke wire technique, as described by Corke et al. (1977) in conjunction with diffuse lighting is being used to visualize the flowfield. For the experiments presented here the smoke wire is positioned vertically approximately 2 mm downstream of the trailing edge of the splitter plate.

Point flow field velocity measurements have been made using standard hot-wire measurements. Multi-sensor hot-wire measurements will also be employed to investigate the instantaneous spatial and temporal coherence of the large scale vortices in the flow. Laser anemometry measurements will be employed to further check the velocity field measurements and to determine if the injected particles significantly alter the flow through possible two way momentum coupling.

The particle dispersion process is being examined using both high speed photographic techniques and laser velocimetry measurements. Lighting for the high speed
photographs employs both diffuse and laser sheet techniques. Views from various planes of the dispersion process will be used to develop a full three-dimensional picture of the dispersion process. Image analysis of the flow visualization and the particle dispersion process is being carried out using a processing system consisting of an Intel Multi-bus with an 8086 microprocessor and imaging technology IP-512 series digitizing and memory boards.
RESULTS

Mean velocity profiles of a $\lambda = 0.2$ mixing layer obtained at several downstream locations are shown in Figure 3a. The measured downstream spreading rate of the mixing layer, $d\delta U/dx = 0.17 \lambda$, where $\delta$ is the local vorticity thickness, agrees closely with previous mixing layer studies (Broward and Troutt, 1985) carried out over a wide range of $\lambda$ values.

Figure 3b shows profiles of downstream turbulence intensity levels for the $\lambda = 0.2$ mixing layer which are in good agreement with previously reported results (Browand and Latigo, 1979).

A smoke flow visualization of the mixing layer flow is shown in Figure 4. The visualization shows the well known large scale vortex nature of the plane mixing layer. Vortex interactions between neighboring vortices are also apparent in the visualization.

Figure 5a shows the vortex structure of the plane wake flow produced when the two free streams have identical velocities, or $\lambda = 0$. The large scale vortex nature of the plane wake is very reminiscent of the well known Von Karman vortex sheet produced by the flow downstream of a circular cylinder. This flow has large spanwise vortex structures of differing signs as compared to the mixing layer which has only spanwise vortices of the same sign. There appears to be less large scale vortex interactions or pairings occurring in this flow because of the stability of the opposite sign vortex arrangement. Figure 5b, however, demonstrates that acoustic forcing at a frequency corresponding to the first subharmonic of the natural vortex passage frequency does have a large effect on the vortex structure development. This result is similar to the observations of Roberts and Roshko (1985) for a plane wake flow in water with subharmonic forcing.

Photographs of the instantaneous dispersion of 41 $\mu$m diameter solid glass spheres of density 2.4 g/cm$^3$ are shown in Figures 6a and 6b. The particles are released near the high speed edge of the mixing layer approximately two boundary layer thicknesses downstream of the plate trailing edge. The mixing layer free streams are at 4 m/sec and 2 m/sec for these experiments.
Figure 3b. Turbulence intensity profiles for $\lambda = 0.2$ mixing layer.

Figure 4. Flow visualization of $\lambda = 0.2$ mixing layer.

Figure 6a gives a closeup of the particle dispersion pattern slightly downstream of the injector. Obviously, the organized vortex patterns have a strong effect on the particle dispersion process in this region. The variation of the $\gamma$ parameter in this photograph is from approximately 25 to 2, and $\gamma$ is defined as the ratio of $\tau_A$ to $\tau_f$.

Figure 6b displays the particle dispersion over a longer downstream length of mixing layer development giving a variation of $\gamma$ from approximately 25 to 0.5. The initially orderly pattern of the particle dispersion appears to break down into a chaotic pattern near the downstream portion of the photograph.

The experimental particle dispersion results are in good qualitative agreement with numerical simulations shown in Figures 7a and b from Chung and Troutt (1987). Although the particle dispersion simulations are for the case of an axisymmetric jet rather than a plane mixing layer, the qualitative nature of the dispersion patterns are quite similar. The $\gamma$ values shown on the simulated dispersions are calculated using the jet diameter as a length scale. The vertical and horizontal axes on the diagram are also nondimensionalized with the jet diameter.
Figure 5a. Flow visualization of wake flow, nonforced.

Figure 5b. Flow visualization of wake flow, forced at first subharmonic of vortex passage frequency.

Figure 6a. Visualization of particle dispersion, close up.
The experimental results obtained from this study demonstrate the importance of large scale vortices in the dispersion of solid particles in free shear flows. Qualitative interpretations of the experimental results also show good agreement with recent numerical simulation results.

The obvious next step in the experimental program is to obtain quantitative information concerning the particle dispersion process. There are two general approaches to obtaining this type of information. The most straightforward approach is to use laser velocimetry techniques to obtain statistical information concerning the velocity and concentration of the dispersing particles at various position throughout the flow. This type of measure-
ment technique has been used successfully in previous particle dispersion experimental studies and a detailed description of its application can be found in Wells (1978).

The second type of approach for quantifying the particle dispersion process involves the use of image analysis techniques to analyze the particle dispersion from high speed photographic results. Several experimental studies have recently been directed towards this type of data analysis using both manual and automated digitizing techniques. A recent review of these techniques and their relative advantages can be found in Altman (1985).

SUMMARY

The experimental results obtained from this research program support the hypothesis that organized vortex structures strongly influence the particle dispersion process. The results also support the idea that particles with intermediate time scale ratios may be most strongly influenced by the organized vortex structures. In addition good qualitative agreement between the present experimental results and recent numerical simulations indicates that advanced modeling techniques may have good potential.

REFERENCES

Altman, D.B., 1985, Laboratory studies of internal gravity wave critical layers, PhD Dissertation, University of California at San Diego.


STABILITY OF THIN FILMS

S. G. Bankoff and S. H. Davis

Northwestern University
Evanston, IL 60201, U.S.A.

ABSTRACT

We consider static liquid layers on planar solid boundaries and analyze their instabilities. The plates are heated and thermocapillary, evaporative and rupture instabilities are discussed with particular attention paid to the development of dry out. Long-wave evolution equations are derived that govern the nonlinear stability of the layers subject to the above coupled mechanisms. Previous theory, experiment and present results are thoroughly discussed.

Liquid layers on planar solid boundaries are susceptible to instabilities that can drastically alter the configurations of the layers.

There are dynamic modes that are generated by the vorticity of the flow [1] and those that couple the flow to the interface deflection [2,3]. Both of these lead to the generation of waves of large amplitude.

There are modes that are present in ultra-thin layers (100-1000Å) that are driven, even in static layers, by long-range molecular forces, van der Waals (vdW) attractions [4] and result in the rupture of the layer. This rupture leads to the layer being broken into dry and wet sections separated by contact lines.

The vdW forces have been modeled [4] via a potential $\phi$,

$$\phi = \frac{A'}{6wh^3}$$

where $h'$ is the local layer thickness and $A'$ is a Hamaker constant; we consider here the case $A' > 0$, which corresponds to a destabilizing vdW force. Static stability theory [5] delivers a condition for the instability of a static layer. A dynamic linear stability theory [6], based on the Navier-Stokes equations with the extra body force $\nabla \phi$, shows that the initial instability, periodic along the bounding plane, has its critical wave length $\lambda_c$ much larger than the mean depth $h_0$ of the layer. The interfacial deflection of this instability can be described by a nonlinear stability theory based on the long-wave nature of the response. Williams and Davis [7] show for $\alpha > 0$,

$$\alpha = \frac{2wh_0}{\lambda},$$

that the interface shape $h = h'/h_0$ satisfies in two dimensions the nonlinear
where \( S \) and \( A \) are dimensionless parameters measuring the surface tension and the Hamaker constant \( A' \), respectively; subscripts denote partial differentiation. Equation (3) regains the results of Ruckenstein and Jain [6] if it is linearized about \( h = 1 \). When equation (3) is integrated numerically for \( x \)-periodic boundary conditions [7], one finds that the layer thickness goes to zero at discrete points in a finite time \( t_R \), the rupture time.

It is clear that the breakdown of the layer into one with dry regions can occur in more complex systems in which case the rupture instability can interact with others. In fact, if one has a relatively thick film, another instability can create locally thin areas in the layer in which the vDW attractions can become important and rupture can occur. Thus, there can be a two-stage process, viz. a dynamical thinning followed by a rupture instability. We pursue this line of inquiry here for a static layer placed on a heated plate.

A liquid layer on a heated plate can be affected by non-uniformities of the temperature activating thermocapillary (TC) effects on the interface. If the plate is heated uniformly and the interface is planar and parallel to the plate, there is no TC effect. It is only when the interface is perturbed into a (say, an \( x \)-periodic) corrugated form that the interface experiences temperature gradients. Davis [8] generalizes equation (3) to this case and finds that \( h \) satisfies in two dimensions the evolution equation

\[
\frac{\partial h}{\partial t} + S(h^3h_{xxx})_x + \{(MP^{-1}h^2 + Ah^{-1} - Gh^3)h_x\}_x = 0
\]

(4)

Here \( P \) is the Prandtl number and \( M \) is a dimensionless parameter, the Marangoni number, measuring the temperature gradient across the layer. The layer is taken to be horizontal with the dimensionless gravity \( G \), the Bond number, acting from the bounding gas toward the liquid. Equation (4) is obtained under the proviso that at leading order in \( \alpha \) the convective heat transported across the layer is smaller than the conduction and that the gas environment is a poor heat conductor. Analysis of this equation by linearization about \( h = 1 \) shows how TC effects are destabilizing in that fluid is pumped from thin to thick regions of the perturbed layer by TC forces.

If the plate is heated non-uniformly, then TC forces create a dynamic state always. Davis [9] has generalized equation (4) for this case and obtains the evolution equation

\[
\frac{\partial h}{\partial t} + S(h^3h_{xxx})_x + \{(M_0P^{-1}h^2 + Ah^{-1} - Gh^3)h_x\}_x - M_1P^{-1}(T'(x)h^2)_x = 0
\]

(5)

Here \( M_0 \) is the Marangoni number measuring the mean vertical temperature gradient while \( M_1 \) measures the horizontal gradients in the plate temperature \( T \), where \( T(x) \) is a slowly-varying function on a scale comparable to \( \alpha \); the prime denotes differentiation. Notice here that \( h=\text{constant} \) is not a solution to equation (5), given that the interface will dimple (if the surface tension decreases with temperature) over portions of the plate where \( T(x) \) is large. We shall report the results [10] of the numerical solutions of equation (5) relating the degree of dimpling to the magnitudes of the parameters and analyzing the appearance of rupture.
instabilities. Figures 1 and 2 give some typical results. Figure 1 shows the steady-state profile for a relatively thick film (> 10⁻⁷ m), such that van der Waals forces may be neglected, under the influence of a horizontal temperature gradient in the support surface. It is seen that surface tension effects are always small, n is produced only if surface tension effects are taken into account. Figure 2 shows the influence of the magnitude of the dimensionless horizontal temperature gradient on the steady film profile. It remains to analyze numerically the instabilities in this deformed state which will lead to the appearance of dry patches. It is straightforward to generalize eqn (5) to three-dimensional disturbances and it is a 3D stability that may in fact be the most dangerous one.

If the plate is heated and the liquid is volatile, then the layer will evaporate, causing the exchanges of mass, energy and momentum. If the process is in quasi-equilibrium so that the interface temperature is constant, TC effects will be absent. However, evaporative instabilities may still be present. Bankoff [11] shows that vapor recoil is a cause of instability. Burelbach, Bankoff and Davis [12] have generalized equation (2) to this case to obtain the evolution equation

$$ h_t + E h^{-1} + S(h^3 h_{xxx})_x + \left\{ (A h^{-1} + E^2 D^{-1}) h_x \right\}_x = 0 \quad . \quad (6) $$

Here E is a dimensionless evaporation parameter that is inversely proportional to the latent heat and D measures the ratio of the vapor to the liquid density. Equation (6) is not satisfied by $h$-is-constant due to the mass loss. The basic state is a thinning static layer, $h = \bar{h}(t)$,

$$ \bar{h} = (1-2Et)^{1/2} \quad (7) $$

which "disappears" in a finite time $t = t^*_D = 1/2E$. The slowly-varying assumptions that lead to equation (6) breakdown near $t = 0$ and $t = t^*_D$; the equation holds in only the (long) excluded interval.

The term proportional to E in equation (6) measures the mass loss during evaporation while that proportional to $E^2$ derives from the vapor recoil. Evaporation is greater at the troughs than at the crests of the waves, so that it is easy to see that the vapor recoil destabilizes the layer. We shall present numerical solutions [12] of equation (6) showing how evaporative and rupture instabilities interact. Further, we shall give extensions of equation (6) to non-equilibrium evaporation in which the temperature of the interface is no longer constant but depends on the fluxes; in this case TC effects are present as well.

In all of the above situations the configurations of liquid layers can be strongly influenced by instabilities. We are particularly interested in the phenomena of dry-out in which dry patches or rivulets are formed from continuous films. Non-isothermal systems are particularly susceptible to such occurrences since TC or evaporative effects create thin regions that may readily rupture.

We have analyzed numerically the time dependent equation and have obtained results for various ranges of parameters. Figures 3 and 4 illustrate the results. Figure 3 shows the development with time of an initially-sinusoidal perturbation of a non-evaporating (isothermal) film. Here the van der Waals forces drive the film to rupture despite the stabilizing effect of surface tension. The dimensionless rupture time of 1.87 is quite close to 1.9 computed by Williams and Davis (1982) for the same problem. Figure 4 shows the time development for an evaporating film with thermocapillary, mass loss, vapor recoil and van der Waals effects present, and
STEADY-STATE FILM PROFILES

$M_j = 0, A = 0, M_j P^{-1} G^{-1} = 1.255$

![Graph](image)

$\xi = kx$

Figure 1. Thermocapillary effects on a thick ($A = 0$) non-evaporating ($E = 0$) liquid film. The effect of the surface tension ($S > 0$) is to produce a smooth solution.

STEADY-STATE FILM PROFILES

$S > 0, M_j = 0, A = 0$

![Graph](image)

$\xi = kx$

Figure 2. Effect of increasing horizontal temperature gradients on the deformation and rupture of a steady-state liquid film.
Figure 3. Unsteady film profiles for isothermal thin liquid film on a horizontal surface. The dimensionless rupture time is in close agreement with that of Williams and Davis (1982).

Figure 4. Unsteady film profile for water with evaporation, mass loss, thermal surface non-equilibrium, surface tension, vapor recoil and van der Waals attractive effects included.
surface thermal equilibrium (K=0). Hence no thermocapillary effects are present. These parameters correspond to a water layer of initial thickness $10^{-1}$ m evaporating into an overlying layer of pure vapor at atmospheric pressure under a temperature difference of 10 K. The dimensionless wavelength of the initial perturbation is 140, so that lubrication theory applies. The dimensionless rupture time is 20.2, which corresponds to $2.02 \times 10^{-6}$ seconds. Hence the rupture is, for all practical purposes, instantaneous.

Equation (6), derived for an evaporating film, simultaneously holds for a condensing film if one takes $E < 0$. In this case the basic state is a uniformly thickening film still subject to vapor recoil and van der Waals instabilities.

This work has been supported by a Grant with the U. S. Department of Energy, Office of Basic Energy Sciences.

References

EXPERIMENTAL AND THEORETICAL STUDIES OF BINARY VAPOR CONDENSATION

M. B. Frish, R. Waterhouse, and G. Wilemski

Physical Sciences Inc., Dascomb Research Park, P.O. Box 3100
Andover, MA 01810

ABSTRACT

This paper describes recent experimental research on the homogeneous nucleation and subsequent growth of binary clusters formed by adiabatically supercooling mixtures of condensable gases in a supersonic nozzle. The status of binary nucleation theory is also briefly reviewed.

INTRODUCTION

Adiabatic expansion in a supersonic nozzle is a convenient way to obtain rapid supercooling of condensable gas mixtures. In studying the condensation of a cooling vapor, determinations of the nucleation rate, critical cluster size, and post-critical droplet growth rate are of prime interest. Direct measurement of these quantities in a supersonic nozzle has so far transcended experimental capabilities. Instead, the desired information is usually extracted by combining directly measurable quantities with gasdynamic relations and theoretical models of condensation processes. [1] This paper presents measurements of binary condensation onset conditions in a supersonic nozzle. Our experimental apparatus combines interferometry, static pressure measurements, and Rayleigh light scattering to provide the most detailed spatial resolution of the onset and droplet growth regions ever obtained. Progress in binary nucleation theory is also briefly discussed, and comparisons between experimental and theoretical onset conditions are made whenever permitted by the availability of key thermodynamic data.

BINARY NUCLEATION THEORY

Theories of binary nucleation have received considerable attention in the past several years. Four theories or calculational methods are presently available. The theories differ in the way the critical nucleus composition is determined and interpreted and, consequently, in the way the free energy of critical nucleus formation is calculated. The earliest theory, that of Reiss [2] and Doyle [3], uses the macroscopic surface tension with two "generalized" (but thermodynamically inconsistent) Kelvin equations that contain compositional derivatives of the surface tension. The Kelvin equations determine the critical nucleus composition and are thus crucial to the evaluation of the critical free energy. This theory is known to be inadequate for describing binary nucleation in water-rich aqueous alcohol or acetone mixtures for which surface enrichment effects occur in bulk liquids. [4] The three newer theories provide different ways of resolving this theoretical inadequacy in relatively successful fashion. Disagreements over the conceptual basis for binary nucleation theory

115
still abound, but the ability to account computationally for experimental results is good, and this provides some measure of confidence in predicting the behavior of systems not yet studied experimentally.

Details of these theories cannot be given in this limited space, but they do share a common feature: the explicit recognition that the macroscopic surface tension intrinsically embodies the effects of surface enrichment occurring in a bulk system. Each of the new theories modifies either the use of the macroscopic surface tension or the way in which the critical composition is determined or both, depending on whether or not surface enrichment is assumed to be present in the nucleating cluster. The approach developed by Flageollet-Daniel, Garnier and Mirabel [5] uses a lattice model to explicitly account for differences between the cluster interior and surface compositions. The cluster surface tension is the macroscopic surface tension of a solution whose composition is equal to that of the cluster interior. The theory developed in the present research program relies only on the macroscopic surface tension but uses thermodynamically consistent Kelvin equations to determine the interior cluster composition, implicitly accounting for surface enrichment. [6,7] The third theory, proposed by Spiegel, Zahoransky and Wittig [8] and Rasmussen [9], assumes that surface enrichment is negligible in the cluster and that the dynamic surface tension of a freshly formed interface should be used. The dynamic surface tension can be computed using the same lattice model cited above, but this approach still requires, either explicitly or implicitly, using the thermodynamically inconsistent Kelvin equations from the original theory to determine the cluster composition. As illustrated in Figure 1, all three theories are in good agreement with experiment at low nucleation rates. However, at much higher rates occurring in shock tubes and nozzles, our purely thermodynamic theory [7] may yield unphysical behavior for highly nonideal systems, whereas the lattice model approaches remain well behaved. The experiments discussed next were designed to obtain the data needed to further test these various theories.

![Figure 1. 1-Propanol/Water Onset Activities at Nucleation Rate of 1 cm\(^{-3}\) s\(^{-1}\)](image)

![Figure 2. Flow System Schematic](image)

**DESCRIPTION OF APPARATUS**

Our experimental apparatus, illustrated in Figure 2, employs two heated plenum chambers to feed an intermittent supersonic nozzle which exhausts into a vacuum. Gas contained in the large (60 gallon) tank is heated to within 5 K of the desired stagnation temperature and contains sufficient volume to maintain this initial temperature to within 0.2% during a 300 ms run. This tank feeds a small (10 liter) plenum which, at the start of a run, contains all of the gas that will flow through the nozzle during that run. The temperature of this gas is maintained to within 1 K of the
The nozzle is defined by two carefully machined nozzle blocks enclosed between two parallel pyrex glass walls. It has an 11:1 contraction upstream of the 0.5 x 1.27 cm sonic throat. The supersonic portion comprises straight, diverging walls with a throat:exit area ratio of 1.36 over its 7.27 cm length. Under usual operating conditions, this design provides cooling rates of less than 1 K/μs and temperatures between 230 and 290 K in the condensation zone located 0.5 to 4 cm downstream of the nozzle throat. Condensible vapors are controllably mixed with the carrier gas and fed directly into the large plenum.

DIAGNOSTIC TECHNIQUES

Pressure Orifice

To quantitatively interpret information derived from the Mach-Zehnder interferometer (discussed below), the density at one point in the flow must be measured by an alternate technique. This is accomplished with a local static pressure sensor comprised of a 450 μm diameter orifice in the upper nozzle block connected to an electrical pressure transducer. The transducer indicates the static pressure at the orifice with a 1/e time constant of about 40 ms. The output voltage of the transducer is recorded and converted to a pressure value by a desktop computer system. The transducer and associated amplifiers were calibrated in situ with a mercury u-tube manometer. There is no evidence of any flow disturbance due to the orifice.

Interferometer

A Mach-Zehnder interferometer is used to determine the local density of the gas in the nozzle. The basic theory of this device is well described in the literature. [10] Its primary feature is that the fringes which comprise the interference pattern when the gas is flowing are shifted in position when compared with those created when the gas is stagnant. Each fringe is usually assigned a number counted from an arbitrary origin within the nozzle. When comparing situations with and without flow, the fringe shift, or change in fringe number at a fixed position can be measured, and this function inverted to yield the density profile within the nozzle.

We have developed an automated system for collecting and analyzing the interference patterns. The system employs a self-scanned photodiode array onto which the interference pattern is imaged. The electrical output signal from the array is fed into the computer, which samples and stores these data twice for each run: once immediately before the flow begins, and again approximately 250 ms later, when the flow has attained steady state conditions. The computer also samples the pressure at these times, allowing automatic computation of the reference density. The data are reduced by measuring the shift in the positions of the fringe minima and maxima when the gas is flowing as opposed to their position during no-flow conditions. The fringe shifts are then converted to density gradients, and the gradients converted to actual densities using the reference density measured at the pressure orifice. The density ratio profile measured by this technique under typical stagnation conditions is shown in Figure 3. Clearly, the agreement with the design density ratio, also shown, is excellent.

The latent heat released when vapors condense within the nozzle causes a deviation from isentropic flow conditions, which can often be detected with the interferometer. By comparing, for example, density ratio profiles obtained under conditions of dry flow to those obtained with condensing vapors, the position within the sampled portion of the nozzle where condensation begins can be located. Figure 4 shows the fractional deviation of the density profile of the dry flow compared to a moist flow, having 0.20 psia of water vapor added to the nitrogen carrier gas with a total stagnation pressure $P_o = 37$ psia, stagnation temperature $T_o = 316$ K, and water mass fraction $w_o = 0.0035$. The width of the curve is equal to the statistical uncertainty in the measurements. Clearly, there is a statistically significant
departure from dry isentropic conditions at a point 1.0 cm downstream of the pressure sensing orifice, indicating the onset of condensation. Using the data of Figure 3 to calculate the effective shape of the nozzle, the moist density profile can be used in conjunction with the diabatic gas dynamic equations to calculate the temperature profile, shown in Figure 5, and the fraction of available vapor which has condensed, shown in Figure 6. From these figures, it appears that detectable condensation begins when the temperature has decreased to 239 K. The water vapor pressure at that point is 3.67 torr, and the saturation ratio is 27.

Rayleigh Scattering

Although the data presented above yield an adequate measure of the conditions at the onset of condensation and can be used to calculate limited information about nucleation and particle growth rates, deviations from the dry flow profile are observable only when the mole fractions of condensible vapors exceed 1 or 2%. Often, the condensible vapors are much more dilute, and the heat released during condensation is imperceptible. To detect the onset of condensation in these situations and gather information about particle growth rates, Rayleigh scattering of laser light is employed.

The principles of Rayleigh scattering are well known. It is sufficient for our purposes to recognize that particles having radii which are small compared to the wavelength of light scatter that light with an intensity that increases as the sixth power of the particle radius. Thus, as condensation nuclei form and increase in size,
the intensity of the light scattered from them increases dramatically. The position of condensation onset can therefore be clearly recognized by measuring the scattered light intensity as a function of position in the nozzle and locating the point at which the intensity begins to increase. To perform this measurement, the 5145A line of an argon-ion laser beam is directed down the axis of the nozzle. Light scattered normal to the beam passes through the nozzle walls and is imaged onto an intensified diode array which is read by the computer.

Figure 7 shows the average relative scattered intensity as a function of position in the nozzle when operating under the conditions of Figure 3. The scattering is seen to be constant upstream of x = 1.02 cm (measured relative to the pressure transducer), while downstream of that position it increases rapidly and ultimately achieves a nearly constant value before the end of the observable portion of the nozzle. The location at which the intensity of scattered light begins to increase corresponds exactly, in this example, to the position at which heat addition to the flow is first observed. Thus, both diagnostic techniques yield the same conditions for the onset of condensation. Furthermore, the combined measurements can be used to calculate particle growth rates. Such calculations will be performed in future work.

![Figure 7. Scattered Light Intensity Profile with Condensation](image)

**EXPERIMENTAL RESULTS AND COMPARISON WITH THEORY**

Condensation onset data have been collected over a range of stagnation conditions for two single component systems, water and ethanol, and for two binary mixtures, water/ethanol and water/propanol. For each system, the total concentration of condensible vapor in nitrogen carrier gas ranged from 0.3 to 4.0 mole percent, with total stagnation pressures between 37 and 45 psia. The experimentally determined vapor pressure and temperature at the onset of condensation for the single component systems, water and ethanol, are shown in Figures 8 and 9 along with representative data gathered by other investigators. Also shown are the equilibrium vapor pressure curves for the condensing species and a sample isentropic P-T trajectory corresponding to the path followed by the vapor in the nozzle. For ethanol, there is very good agreement between our results and those obtained by Wegener et al. in nozzle experiments [1] and those obtained by Peters in shock tube experiments [11]. For water condensation, the question of agreement is a little more difficult to resolve because there are considerable differences between the sets of data reported by other investigators as shown. Our results compare very well to those of Barschdorff [12] which were obtained in shock tube experiments, but show condensation to occur at higher temperatures (for the same initial water vapor pressure) than those measured by Wegener and Pouring [13] and by Jaeger et al. [14] in nozzle flows.
The discrepancies may be explained by the differences in the cooling rates achieved in the different experimental techniques, along with the differing sensitivities in the techniques used to locate the onset of condensation. For instance, Barschdorff located onset by looking at scattering with a 15 mW HeNe laser, while we use a 100 mW argon ion laser. The combination of greater intensity and shorter wavelength from our laser would be expected to yield greater sensitivity to smaller droplets. The cooling rates achieved in our nozzle are between 300 and 400 K/ms, while those reported for shock tubes range from 50 K/ms [8] to 120 K/ms [12], and other nozzles range from 600 K/ms [13] to 2500 K/ms [14]. On this basis alone, the position of our data between the data from Wegener's nozzle and that from Barschdorff's shock tube could be expected. (However, it has been reported [11,15] that, for some systems, the effect of cooling rate on the adiabatic supercooling achieved is very small.)

Due to the added variable of mixture composition in addition to temperature and pressure, the data for the condensation of binary mixtures cannot be simply represented on a single P-T diagram. To facilitate analysis, our data for the binary systems were acquired by fixing the initial temperature, pressure, and total vapor pressure of the condensable species so that all of the data for one system fell along a single isentrope. As the relative concentrations of water to alcohol in the mixture were varied, the location of the onset of condensation would move up or down this isentrope. Using this single-isentrope frame of reference, we can plot the temperature at onset as a function of the composition of the binary mixture. The results are shown in Figure 10. The initial conditions used for the water/ethanol system were: \( T_0 = 316 \text{ K}, P_0 = 37.0 \text{ psia} \), and total vapor concentration = 0.50 mole percent. For the water/propanol system: \( T_0 = 343 \text{ K}, P_0 = 41.7 \text{ psia} \), and total vapor concentration = 0.60 mole percent. While the temperature, pressure, and total vapor concentration for the water/propanol system are higher than for the water/ethanol system, they actually lie along the same isentrope, making possible a direct comparison in the same figure. The endpoints of the two sets of data in Figure 10 represent the single component onset temperatures for water, ethanol, and propanol along this isentrope. The shape of the curve between these endpoints is determined by the properties of the mixture, including vapor pressure, density, and surface tension and should be accessible by nucleation theory. Unfortunately, the comparison of these experimental points to theoretical predictions is currently hindered by the absence of the essential thermodynamic data at low temperatures.
Another popular means of portraying condensation data for binary mixtures is to plot the activity of the alcohol component vs that of the water component for the onset of condensation at a single temperature. (Activity is defined as the quotient of the partial pressure of the vapor component at the onset of condensation and the equilibrium vapor pressure of that component at the same temperature.) Activity plots, based on the data shown in Figure 10, are shown in Figures 11 and 12 for the water/ethanol and the water/propanol systems respectively. These differ from the usual activity plots in that the points are not strictly isothermal but correspond to decreasing temperatures going from low to high water activity.

Figure 12 also compares the data with several theoretical curves for binary nucleation onset of 1-propanol/water mixtures at constant temperature and nucleation rate. Thermodynamic properties at 253.2 and 263.2 K, where deficient, were obtained by extrapolating from higher temperatures, and solution activity coefficients were computed from a higher temperature fit, so these comparisons are by no means definitive. The curves were calculated using the dynamic surface tension approach which, despite its thermodynamic inconsistency, has proven successful in correlating experimental onset measurements from diffusion cloud chambers and shock tubes. Below a water activity of three, our experimental results are reasonably well represented by
the three theoretical curves at 273.2, 263.2, and 253.2 K with a nucleation rate of $10^{11} \text{ cm}^{-3} \text{ s}^{-1}$. This rate is the same as has been found for shock tube experiments, which report cooling rates somewhat lower than ours. The theoretical curve at 253.2 K, for a rate of $10^{16} \text{ cm}^{-3} \text{ s}^{-1}$ gives an indication of how much the predicted onset conditions vary with nucleation rate. Based on this limited comparison, the lower rate curves seem to be more appropriate for correlating our results. This conclusion needs to be strengthened by obtaining more data at higher water activities and higher temperatures since the few data we now have at high water activities are at temperatures too low for reliable theoretical comparisons. The same difficulty exists for all of our ethanol/water data. We are currently working to obtain the requisite high temperature measurements.

ACKNOWLEDGMENT

This work was supported by the United States Department of Energy, Division of Engineering, and Geosciences, Office of Basic Energy Sciences.

REFERENCES

EXPERIMENTS IN AUTONOMOUS ROBOTICS*

William R. Hamel
Oak Ridge National Laboratory
Center for Engineering Systems Advanced Research
Oak Ridge, Tennessee, USA 37831-6007

ABSTRACT

The Center for Engineering Systems Advanced Research (CESAR) is performing basic research in autonomous robotics for energy-related applications in hazardous environments. The CESAR research agenda includes a strong experimental component to assure practical evaluation of new concepts and theories. An evolutionary sequence of mobile research robots has been planned to support research in robot navigation, world sensing, and object manipulation. A number of experiments have been performed in studying robot navigation and path planning with planar sonar sensing. Future experiments will address more complex tasks involving three-dimensional sensing, dexterous manipulation, and human-scale operations.

INTRODUCTION

The Center for Engineering Systems Advanced Research (CESAR) was established at Oak Ridge National Laboratory in 1983 as a national center for multidisciplinary, long-range research and development in machine intelligence and advanced control theory for energy-related applications. Intelligent machines are operational systems capable of autonomous decision making and action. Initial research emphasis is on remote operations, with specific application to dexterous manipulation in unstructured, dangerous environments. Potential benefits include reduced risk to man, machine replication of scarce expertise, minimization of human error due to monotony and fatigue, and enhanced capabilities through sensors and computers. A CESAR goal is to develop the interface between today's advanced teleoperation and the autonomous machines of the future.

CESAR was created by the Division of Engineering and Geosciences, which is part of the Office of Basic Energy Sciences in the U.S. Department of Energy. The initial CESAR research objectives and approach [1] have evolved with time [2-4] and are currently documented in a five-year plan [5] which is updated annually. At this time, CESAR research concentrates on the issues related to autonomous robots for hazardous and unstructured work environments. Research activities include development of methods for real-time route planning with sensor feedback, determination of concurrent algorithms for optimal implementation on advanced parallel computers, formulation of learning methodologies for knowledge acquisition and interpretation, uncertainty modeling and analysis, machine vision based on human ocular processing, and compliant manipulator dynamics and control. The initial phase of CESAR research has been performed using a small-scale mobile robot that has many of the essential functional attributes of an autonomous robot.

The purpose of this paper is to discuss the important role of experimentation in robotics research and to describe the types of experiments and experimental systems being incorporated into CESAR's activities. Future experimental plans are also presented.

ROLE OF EXPERIMENTATION

CESAR's planning foundation and research agenda are based upon the premise that experimental validation and demonstration of robotics concepts is essential. Autonomous robotics research involves development of theories and concepts involving virtually every discipline of science and engineering. Such research is also tightly coupled to the engineering practicalities of actual robot fabrication--implementation of theories and concepts is constrained by what can be designed and built.

Autonomous robot experimentation provides mechanisms for demonstrating and evaluating system concepts and principles. Because the goal of robotics research is to advance man's ability to "build" intelligent robotic machines that more greatly benefit society. Such ideas, concepts, and theories are rightfully judged by their performance in experiments. Robotics experiments provide proof-of-principle evaluations of theories and algorithms. Intelligent machines and robots must not only emulate intelligent behavior, they must also be capable of sensing and responding to influences of their environment. Consequently, robot sensors are a major component of autonomous robotics research. Robot sensors are not unlike sensors and transducers in general except for major requirements and interests in vision sensing and interpretation. As in humans, vision is the predominant sensing channel for intelligent mobile robots. Robot sensor development requires basic experimental research in the way in which the physical transduction phenomena of interest are embodied and the performance of the associated sensors is evaluated. An important side benefit of robotics experimentation is the feedback it provides to the engineering implementation of advanced concepts. The exercise of attempting to implement specific concepts provides engineers with an opportunity to study and evaluate robot system implementation research issues. Such basic issues are particularly apparent in mechanization of robotic functions and in the computation hardware and software necessary to implement complex real-time algorithms.

Finally, as in all experimentation, robotics experiments are a powerful vehicle for uncovering new research issues and for categorizing the significance of known research issues. Effective robotics research must properly balance theoretical development with experimental assessments.

EXPERIMENTAL PARADIGM

CESAR is dedicated to both the theoretical development of advanced autonomous robot concepts and the experimental evaluation of such concepts. It is recognized that many, if not all, of the research challenges in autonomous robotics will derive from real-world constraints and practicalities. Useful robots must be able to perform in some sense what human workers can do under typical environmental conditions. This "goal" has many ramifications, but two obvious ones are (1) the robot's sensors, especially vision, must function under realistic conditions (e.g., lighting cannot be overly manipulated or contrived), and (2) robot manipulators and other resources must be able to manipulate objects and tools that are at least within the human range of force. It is believed that defining the experimental environment on a human scale is a significant research factor (in terms of driving objectives). Because of this, and to give general context to our long-range research planning, the CESAR team established [5] a reference task problem, or paradigm, about which to organize our goals and objective. The general problem is the operation, diagnosis, and maintenance of process control valves. The ultimate goal is development of an autonomous mobile robot capable of repairing or replacing typical process valves under less than ideal (perhaps emergency) conditions. The process valve problem was selected for several reasons: (1) valves are very common in energy-related systems; (2) valve operation,
diagnosis, and maintenance tasks cover a very wide range of complexity, load range, and force-sensitivity requirements; and (3) typical valve installations in real plants provide difficult mobility, manipulation, and sensing challenges (refer to ref. 5 for more detail of the rationale). It is believed that a robot capable of accomplishing these representative tasks would inherently be capable of performing a wide range of typical human tasks.

EXPERIMENTAL SYSTEMS

The initial experimental focus of CESAR has been a mobile robot system called HERMIES-II [4] (Hostile Environment Robotic Machine Intelligence Experiments: Series II). This robot is a low-cost system developed for basic experiments in autonomous robotics with dual-arm manipulators, on-board distributed digital processors, and a directionally controlled sensor platform. HERMIES-IIA, shown in Fig. 1, is propelled by a dual set of wheels on a common axle alignment and independently driven by individual dc motors. The on-board IBM-PC and other electronics are located in the enclosure above the drive chassis, and the dual-arm manipulator with shoulder torso is positioned immediately above the electronics. The manipulators are recognizable as Zenith/Heathkit HERO robot arms. The sensor platform has dc servocontrolled pan and tilt mechanisms to position a five-axis sonar ring and a pointable combination of sonar sensors and computer vision CCD cameras. The vision system is an International Robomation/Intelligence P-256 unit, which requires tethered operations and provides 256 X 256 pixel resolution with 8 bits of gray level. All HERMIES-IIA control software has been written in the FORTH language. Currently, sonar scan data are preprocessed on-board HERMIES and are transmitted via a 2400-baud RS-232 radio link to either the CFSAR's NCUBE hypercube parallel computer or the LMI Lambda symbolic processor for navigation planning.

HERMIES-IIA has now been upgraded to a new form designated HERMIES-IIB. This upgrade has involved improvements to the robot's basic mobility chassis and on-board computational resources. These modifications have improved reliability as well as increased the degree of "self-contained" autonomy (i.e., dependence on off-board immobile computers has been reduced by using the VME and IBM-AT backplanes in combination). The VME system provides all control and sensor data interfacing. The VME system also serves as a data gateway to the AT backplane, which houses a 4th order (16 nodes) hypercube parallel computer based upon the NCUBE Corp. 32-bit node processor chip. The hypercube host processor, which is an Intel 80286, also serves as the basic robot control engine. The on-board hypercube provides a processing speed equivalent to ~24 VAX/Ill-780s and will be used to implement world models, sensor-processing algorithms (e.g., vision), and navigation/manipulation algorithms. The VME system facilitates on-board integration of a reasonably high-performance computer vision system using DataCube Corp. expansion boards, which provide 512 x 512 x 8 color resolution and traditional image-processing functions. It is believed that HERMIES-IIB represents one of the most computationally powerful mobile robots in existence today.

Robot manipulator dynamics and control constitutes a major CESAR research area. Manipulators are essential for mobile autonomous robots capable of performing tasks typically performed by humans in hazardous environments. It is believed that such manipulators must be efficient, light weight, dexterous, and sensitive in order to perform complex tasks. There is a wide spectrum of research challenges associated with manipulator design, performance optimization, and control.

The CESAR research manipulator (CESARm) shown in Fig. 2 is being developed as a test bed for advanced manipulator design and control methodologies suitable for mobile applications in unstructured environments [4,6]. The manipulator's design was based partly on the design of the master controller for the ORNL Advanced Servomanipulator (ASM [7]). Unlike the master controller, which is fitted with a hand grip, the arm has a parallel jaw gripper. In addition, it has a seventh degree of freedom, which is
desirable for obstacle avoidance and expanding the range of motion. The drive train has low friction and is backdrivable, which should enhance force control. The spherical three-degree-of-freedom wrist, developed at ORNL for the ASM master, has all singularities at the extremities of motion which simplifies the kinematic calculations.

The control system for the CESAR research manipulator is a new design based on the VMEbus. It utilizes Motorola MC68020 microprocessors and the Microware OS-9 operating system. Amplifier and sensor interface electronics are designed for expansion for a second manipulator and future use in a mobile system. The manipulator can exert ~13 kg of load and has a no-load maximum tip speed of ~3 m/s. The CESAR has been fabricated, assembled, and checked out. Presently, advanced position, force, and redundancy control algorithms are being developed and programmed.

CESAR is actively pursing the next phase of research, in which the scale of operations will be increased into the realm of human sizes in terms of manipulation geometry and loads. To accomplish this the HERMIES-III robot is being designed and fabricated. HERMIES-III will be an electric-powered robot incorporating the CESAR research manipulator and a shoulder/torso mechanism mounted on a modified, automatic-guided industrial vehicle (AGV) chassis. Initially, only a single manipulator will be installed, but provisions for adding a second arm at a later date are included.
HERMIES-III will utilize the combined VME/68020 and hypercube computer assembly similar to HERMIES-IIB. A new, faster, and more rigid pan/tilt sensor platform has been developed, and the sensor suite includes the five-element sonar ring, a DataCube CCD camera pair, and a laser-range scanning system. In this configuration, HERMIES-III will be capable of handling relatively large loads from floor level to -2 m off the floor. HERMIES-III will have 3D optical scanning capability and conventional sonar ranging with additional 2D TV scanning. HERMIES-III is in the design phase, although procurement of the laser-range scanner and basic AGV chassis have been initiated.

EXPERIMENTAL ACTIVITIES AND PLANS

The HERMIES-II mobile research robot has been used to perform a number of types of navigation and path-planning experiments. All of these experiments have used a simplified world model consisting of rectangular objects in a flat floor area. HERMIES has been used in four distinct classes of navigation experiments. The first
class was navigation in a fixed environment with precise and complete a priori knowledge of rectangular obstacles. This case is essentially a path optimization problem in which the navigation routine selects a trajectory from the initial condition to the goal state, which is optimal in some sense (e.g., follow the shortest path). In the second class, the a priori world model requirement was relieved, requiring HERMIES to use his sonar scanners to construct an image of the environment as he progresses toward the goal state. Figure 3 is the result of a typical sonar scan, and shows the wide beam width and specular reflection characteristics of sonar that give the robot an extremely blurred perception of his environment. These error characteristics must be incorporated into the path-planning process to minimize the possibility of collisions with obstacles or boundaries. In the third class of experiments, the fixed environmental constraint was also relieved, forcing HERMIES to assume that his perception of the world could change at any time. This generalization forces the robot to "slow down" further and to perform sonar scans more often to be sure that the world model in the robot's vicinity is appropriate for planning.

The most recent (fourth) class [8] of navigation experiments involved the incorporation of real-time collision avoidance and path re-planning. A collision monitoring sonar scans the frontal area of the robot's path to detect unexpected obstacles. In the event an obstacle is detected, the robot controller interrupts the overall navigator routine (implemented on the LMI Lambda computer) and invokes a simple diagnostic routine to evaluate the size and motion characteristics of the obstacle and to implement appropriate recovery plans and paths.

Experiments are also being performed in association with the CESArm. The ultimate implementation, test, and evaluation of advanced manipulator control algorithms requires that a detailed dynamic model of the seven-degrees-of-freedom manipulator be developed. A comprehensive activity is presently under way to quantify the mass properties of the CESArm. A dual approach is being taken involving experimental evaluation and computer-aided design model estimation. The inertial tensor and centers of mass of individual CESArm components are being measured using an

---

![Figure 3. Robot Navigation With Planar Sonar](image-url)
inverted pendulum mechanism in which natural frequency varies with the part position and orientation. Current results suggest that a repeatability of about 3% can be achieved for the simpler shaped components. Separate experiments to calibrate the friction and backlash characteristics of each axis of motion are also under way.

A number of new research activities are now being planned. Initially, HERMIES-IIB will be used to replicate earlier navigation and path-planning experiments [5,6] with full autonomy. Subsequently, a new set of experiments involving combined manipulation and mobility will be performed. In these experiments, HERMIES-IIB will use on-board vision and an optical guidance/control scheme for manipulator positioning to operate a "simulated" process control panel (see Fig. 4). The process control panel will consist of two analog readout meters, two slide-type analog input adjustments, and four back-lighted pushbuttons. The discrete logic and continuous dynamic models that interconnect and drive these inputs and outputs will be implemented with an IBM-PC. Upon finding and establishing position reference with respect to the panel, HERMIES-IIB will "operate" the panel to establish the system output states specified in his original task goals.

Hermies-IIB will use machine vision rather than sonar as its primary sensor. The dual on-board cameras will be used in both 2D and 3D scene analysis for world model updating and will be used as a basis for quantifying robot spatial position commands. A separate experimental setup is being developed for evaluation of vision algorithms. This bench facility will allow moving scene simulations in which objects may have up to three linear axes of precisely controlled motion. Algorithm performance and overall repeatability assessments will be performed under these controlled conditions prior to installation on the robot.

In the longer term, the HERMIES-III robot will facilitate three-dimensional world modeling and sensing [9] and execution of activities involving human-scale tasks and dexterous manipulation. Three-dimensional sensing will be accomplished using a scanning laser radar system.

![Figure 4. HERMIES Process Control Panel Experiment](image)
SUMMARY

CESAR is performing basic research in autonomous robotics for hazardous environments associated with energy-related systems. Experimentation plays a critical role in this type of research by providing a fundamental mechanism for evaluating robot concepts to provide the principles of new theories and algorithms, to study robot implementation issues, and to uncover new research challenges. An evolutionary sequence of mobile robot systems and advanced computers has been developed and used in robot navigation and path-planning research. Future work will focus on machine vision, combined mobility and manipulation tasks, three-dimensional sensing and modeling, and human-scale dexterous manipulation. Experimental evaluation of new concepts will continue to be a key theme of CESAR research.

ACKNOWLEDGEMENTS

In this paper, the author in many respects speaks on behalf of the overall CESAR research staff in that the topics discussed touch on all of their activities. Their cooperation and assistance in preparing this paper are greatly appreciated. Research funding for CESAR is provided by the U.S. Department of Energy (Office of Basic Energy Sciences). The author and the CESAR staff acknowledge the support of A. Zucker, F. C. Maienschein, B. C. Eads, and C. R. Weisbin of the Oak Ridge National Laboratory and O. Manley of the U.S. Department of Energy.

REFERENCES


AN AUTOMATED MULTIVIEWING ULTRASONIC TECHNIQUE FOR FLAW RECONSTRUCTION

D. O. Thompson, D. K. Hsu, and S. J. Wormley
Ames Laboratory, USDOE
Iowa State University
Ames, IA 50011

ABSTRACT

This paper describes key achievements in the development of an automated multiviewing ultrasonic instrumentation for flaw characterization in materials. Based on ultrasonic signals scattered from the flaw in a number of directions, the signal processing algorithms provide a reconstruction of the size, shape, and orientation of the flaw. To improve the reconstruction reliability for arbitrarily oriented flaws, an angular scan plan is developed to select the most favorable data acquisition aperture configuration for a given flaw. Reconstruction results are presented. Electronic circuits are developed for the generation and detection of broadband unipolar ultrasonic pulses in order to characterize flaws over a wider size range. A second generation multiview transducer has been developed to incorporate these capabilities. The instrumentation and the associated signal processing algorithms form a unique research and inspection tool.

INTRODUCTION

The detection and characterization of failure-initiating flaws is an important capability in the assurance of structural integrity in materials and components. Ultrasonically generated elastic waves in the long and intermediate wavelength regime are widely used in the nondestructive evaluation (NDE) of flaw in structural materials. This wavelength regime is characterized by $ka \approx 1$ where $k$ is the wave number and $a$ is the flaw size. Significant theoretical advances have been made in recent years in both direct and inverse scattering in this regime based on physical models that describe the details of the elastic wave-flaw interaction. It has been shown that sufficient information about a flaw (size, shape, orientation and material properties) can be extracted from the scattered ultrasonic signals in this wavelength regime. Such flaw information serves as input needed to implement a failure prediction methodology for materials and components. The physical model-based reconstruction technique places more demanding requirements on ultrasonic instrumentation and signal processing than conventional ultrasonic NDE. These requirements stem from the utilization of absolute flaw scattering amplitudes (both magnitude and phase) and impulse response functions rather than magnitudes of the echoes and transit times.

This project was initiated to develop an automated multiviewing ultrasonic instrumentation and the associated signal processing algorithms for flaw reconstruction and characterization to utilize the results of the inverse scattering theories in the long and intermediate wavelength regime. The design is based on a multiviewing concept in which the flaw is "sized" from a number of interrogation directions to facilitate a 3D reconstruction of the flaw. In the first phase of the project a detailed study of the reconstruction reliability was carried
out, in which it was found that the accuracy of the reconstruction depended strongly on the orientation that a flaw assumed with respect to the aperture of the multiview transducer. It is therefore desirable to have a capability of optimizing the data acquisition aperture based on preliminary estimates of the flaw orientation. To this end, an angular scan plan is developed to improve the reliability for reconstructing arbitrarily oriented flaws.

The utilization of the inverse scattering solutions in flaw sizing requires a broad frequency bandwidth. This bandwidth requirement is often not met by commercial transducers and pulser/receivers. An effort to broaden the frequency response and the ability to accommodate a wider range of flaw sizes has led to the development of generation and detection of broadband unipolar ultrasonic pulses for pulse-echo as well as pitch-catch modes. Developments of the angular scan method mentioned above and the unipolar pulse technique have prompted the development of the second generation multiviewing ultrasonic apparatus.

This paper will review some of the key achievements made in the last three years. The angular scan plan will be described and reconstruction results of arbitrarily oriented flaws presented. The unipolar generation and detection technique, particularly for the pulse-echo mode, will be discussed. Instrumentation development of the second generation multiview transducer will be described and future directions of research are given.

RECONSTRUCTION OF ARBITRARILY ORIENTED FLAWS

In the first generation multiview apparatus the axis of the aperture cone formed by the perimeter transducers is oriented perpendicular to the surface of a flat sample. The effect of adverse aperture conditions on the reconstruction reliability was investigated in extensive computer simulation of the reconstruction. Errors in the reconstruction results were found to increase with increasing tilt angle of the flaw and the degradation was more serious for smaller apertures containing the same number of viewing directions. Experimental reconstruction of tilted flaws were consistent with the simulation results. The accuracy can be restored if the orientation of the flaw is estimated in a preliminary manner and the interrogation aperture is tilted in such a way as to compensate for the flaw orientation. An angular scan plan is therefore developed as a preamble to the actual scattering data acquisition and reconstruction for the purpose of optimizing the orientation of the multiviewing aperture and thereby assuring the reconstruction reliability.

1. Angular Scan Plan

The angular scan plan is designed to provide a rapid but preliminary estimate of the flaw shape and orientation using only the peak-to-peak amplitude of the scattered signals at various azimuthal (a) and polar (b) angles. It is based on the premise that the magnitude of the front surface delta function of the impulse response function is determined essentially by the curvature of the flaw surface at the tangent contact with the wave front. Analogies are drawn with geometrical optics and the backscattered echo amplitude is assumed to be proportional to \( \sqrt{\rho_1\rho_2} \) where \( \rho_1 \) and \( \rho_2 \) are the principal radii of curvature at the contact point on the flaw surface. In order to compute this amplitude as a function of angle, the flaw is modeled as a general ellipsoid with semi-axes \( a_x, a_y, \) and \( a_z \) and Euler angles \( \theta, \phi, \) and \( \psi. \) It then follows that \( \sqrt{\rho_1\rho_2} = a_xa_ayz/r_e \) where \( r_e \) is the flaw center to tangent plane distance for the scattering direction. The value of \( r_e \) depends on the sizes of the semiaxes as well as the orientation of the flaw.

To illustrate the flaw shape and orientation estimation, we computed \( a_xa_ayz/r_e^2 \) for a prolate spheroid with \( a_x=a_z=80\mu m, a_y=250\mu m \) and with its long
axis pointed at an azimuthal angle of 120° and a polar angle of 45° in the laboratory axes fixed on the sample. This represents a short needle-shaped flaw tilted 45° below the solid surface. Figure 1 shows azimuthal scans at five polar angles where the distance from the origin to a point on the curve represents the amplitude of the flaw signal. As can be seen, a plane of mirror symmetry exists at an azimuthal angle of 120° (or 300°). A polar scan at this azimuthal angle, shown in Fig. 2(a), shows a peak at $\alpha=45°$ and $\beta=300°$; thus revealing the tilt angle of the flaw. A second scan in a plane perpendicular to the symmetry plane and containing the direction of maximum signal ($\alpha=45°$, $\beta=300°$), shown in Fig. 2(b), shows a constant signal amplitude and hence confirms the prolate spheroidal shape of the flaw.

![Diagram of azimuthal scan](image)

Fig. 1. Computed flaw echo amplitude in azimuthal scans at five different polar angles. The flaw is a prolate spheroid with semiaxes $a_x=a_z=80$, $a_y=250$ and Euler angles $\Theta=45°$, $\Phi=30°$ and $\Psi=0°$. It represents a short needle-shaped flaw tilted to an angle of 45°.

The angular scan plan therefore consists of two steps. First, the flaw signal amplitude is recorded in azimuthal scans at several (e.g., four) polar angles. From the resulting pattern, a symmetry plane is identified. A polar scan is then made in the symmetry plane to locate the maximum signal direction. The aperture of the multiviewing transducer assembly is then centered about this direction for acquiring a series of pulse-echo and pitch-catch scattering data for flaw reconstruction. Approximate flaw shape information revealed by a third (optional) scan perpendicular to the symmetry plane and passing through the maximum signal direction can then be used as a check of the reconstruction or as a constraint in the regression process. The scan plan was verified experimentally for various flaw geometries.12

2. Reconstruction Results

By using a data acquisition aperture centered about the direction of maximum scattering signal, good reconstruction results were obtained for two tilted flaws. The first flaw is a 400 x 200μm oblate void tilted 30° in a diffusion bonded titanium sample. The second flaw is a short copper wire inclusion with $a_x=a_z=80μm$ and
Fig. 2. Computed flaw echo amplitude of the flaw in Fig. 1 in (a) polar scan at an azimuthal angle of 120° (or 300°) in the vertical sagittal plane (VSP) and (b) scan in the perpendicular sagittal plane (PSP).

$\alpha_y=200\text{um}$ that is embedded at a 45° tilt angle in a thermoplastic disk. Table 1 and Fig. 3 show that the reconstruction accuracy is quite good. It should be noted that poor results were obtained on these two tilted flaws when the aperture axis of the transducer assembly was perpendicular to the part surface.

<table>
<thead>
<tr>
<th>Flaw parameters</th>
<th>Oblate void in titanium</th>
<th>Copper wire inclusion in thermoplastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_x$</td>
<td>394\text{um}</td>
<td>257\text{um}</td>
</tr>
<tr>
<td>$a_y$</td>
<td>419</td>
<td>87</td>
</tr>
<tr>
<td>$a_z$</td>
<td>191</td>
<td>81</td>
</tr>
<tr>
<td>Tilt angle</td>
<td>31°</td>
<td>42°</td>
</tr>
</tbody>
</table>

Table 1. Reconstruction of tilted flaws using apertures determined by the angular scan plan

The application of the inverse scattering solutions in flaw sizing and reconstruction can be enhanced significantly by the utilization of wide bandwidth ultrasonic instrumentation to properly implement the inversion algorithms over a range of flaw sizes. The utilization of unipolar pulses is an attractive way to achieve broad transducer bandwidths. It is known that a unipolar ultrasonic pulse may be generated by a step excitation voltage. The generation requires the transducer to be driven by a low impedance pulser whereas the detection requires the transducer to be connected to a high impedance receiver. For this reason, the application of unipolar pulses has been limited in the past to the pitch-catch mode only. Driven by the needs for broad transducer bandwidth, an effort in this project has resulted in a new circuit design that allows the generation and detection of unipolar stress pulses in both the pulse-echo and pitch-catch modes for the first time. The circuit is a transmit/receive switch that connects the transducer electronically to a low impedance step function pulser during the transmit cycle.
and tilted 45° and the reflected...}

Fig. 3. Ellipsoidal reconstruction (dashed line) and the actual size and shape (solid line) of a short copper wire inclusion with $a_x=a_z=80\mu m$, $a_y=200\mu m$ and tilted 45°.

and to a high impedance receiver during the receive cycle. In this circuit, the parameter $R_1C_\omega \gg 1$ and the received signal does not suffer differentiation effects at the receiver. Here $R_1$ is the impedance of the external circuit as viewed by the transducer, and $C_\omega$ is the reciprocal of the transducer source impedance.

The circuit developed has been used to generate unipolar pulses with standard, commercially available transducers. Using a step excitation voltage, unipolar pulses were generated from a 10MHz, 1/4 in. immersion transducer and the reflected pulses from a planar reflector at different distances were detected using the same transducer. Figure 4 shows the received pulse shapes for different propagation distances. As the propagation distance increases, the pulse shape approaches that of a bipolar pulse because of the preferential loss of the low frequency signal content due to diffraction. The unipolar pulse circuit has now been incorporated into the multiviewing apparatus for improved bandwidths.

One of the tests that must be applied to the extended bandwidth is to compare the measured scattering amplitude with theoretical predictions for a well-defined ultrasonic target. Such a comparison is shown in Fig. 5 for the oblate void in titanium. The solid curve shows the experimentally measured absolute scattering amplitude using the unipolar pulse in the pulse-echo mode and the dashed curve is the calculated result for this flaw.16 These curves are given in absolute scale and diffraction and attenuation effects have been corrected. The agreement is considered very satisfactory and would not have been possible had the bandwidth been narrow.

**MULTIVIEWING ULTRASONIC INSTRUMENT--SECOND GENERATION**

The first generation of the multiviewing transducer assembly consisted of six peripheral transducers and a center transducer. The six transducers were held in a gimbled mount on a 2" diameter ring. They were coupled to a common drive so that all six transducers were set to the same angle of incidence, up to 30° in water. As a result, the instrument was limited to a configuration in
Fig. 4. A unipolar ultrasonic pulse reflected from a planar reflector at different distances from the transducer.

Fig. 5. Comparison of the experimental (solid curve) and theoretical (dashed curve) absolute scattering amplitude of the 400 x 200µm oblate void in titanium.

which the axis of the viewing aperture must be perpendicular to the surface of the part in order for all transducers to "focus" onto the flaw. This design has now been substantially extended and improved in the second generation model completed recently. In the new instrument, shown in Fig. 6, a total of six transducers are used, the fixed diameter mounting ring is eliminated and each transducer can move
radially up to 5" from the center axis. Two of the transducers can be moved to the center in the search mode to serve the function of the central transducer in the previous design. The increased overall diameter of the apparatus makes possible the evaluation of deep flaws up to a depth of 4" in common metals. Each transducer can be tilted to a maximum angle of 60° and translated along its axis over a distance of 2". The three independent degrees of freedom of the six transducers, together with the x, y, z, and θ motions of the whole apparatus, give the multiviewing system a total of 22 versatile adjustments. The new apparatus is therefore capable of carrying out the angular scan plan described above and thereby optimizing the data acquisition aperture configuration for any arbitrarily oriented flaw it encounters.

All the various motions and the positional and angular settings of the transducers are controlled and monitored by a computer. The precision is vastly improved by the use of LVDT encoders for each degree of freedom. The transitional motions are accurate to about one tenth of a wavelength at 20MHz in metals and the angular settings are accurate to ±0.05°.

The transducers are a matched set of commercial immersion probes with a nominal center frequency of 10MHz and a diameter of 1/4". The spread in their spectral amplitudes is within 2-3dB. The transducers can be driven by either a spike voltage pulse or a step function voltage (for unipolar pulse generation) and the various pulse-echo and pitch-catch combinations are selected through a high speed electronic switch. The detected flaw signal is transmitted to a common receiver through 20dB preamplifiers (and the special unipolar pulse circuitry if a step voltage is used) connected to each of the transducers. The time domain flaw signals are then digitized and averaged in a Tektronix 7912 programmable digitizer, and stored in a MicroVax computer for further processing. The signals are processed through a measurement model algorithm to obtain the absolute scattering amplitude and the flaw sizes (center to tangent plane distances) in the various scattering directions are obtained by using the 1-D inverse Born algorithm. Details of these procedures were given in earlier publications.

---

Fig. 6. Photographs at two perspectives of the second generation multiviewing transducer assembly in an immersion tank.
SUMMARY

Significant advances have been made in the second phase of this project toward the engineering development of a quantitative ultrasonic flaw characterization technology based on inverse scattering theories in the long and intermediate wavelength regime. The reconstruction reliability for flaws of unknown orientation has been improved considerably by optimizing the multiview aperture configuration using the angular scan plan. The ability to size flaws over a wider range has been improved with the use of broadband unipolar ultrasonic pulses. The extremely versatile electromechanical system for transducer control and the extensive signal processing software package makes the second generation instrumentation a unique research and inspection tool for ultrasonic NDE. With the automated multiviewing ultrasonic system at hand, plans have been made to investigate problems important to materials reliability. These include the development of 2D and 3D tomographic imaging techniques, characterization of material property gradients and interfaces, and applications to flaws in anisotropic materials and composites.

ACKNOWLEDGEMENT

The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-ENG-82. This work was supported by the Director of Energy Research, Office of Basic Energy Sciences.

REFERENCES

HIGH-SENSITIVITY OPTICAL PROBE FOR PHOTOTHERMAL MEASUREMENTS

J. T. Fanton and G. S. Kino
Edward L. Ginzton Laboratory
W. W. Hansen Laboratories of Physics
Stanford University
Stanford, California 94305

ABSTRACT

A new displacement sensor for photothermal and photoacoustic measurements has been developed. This sensor employs a Wollaston prism to split a laser beam into two components which are focused to the sample where they interferometrically measure the periodic thermal expansion of the surface. Because the two beam paths are nearly coincident and equal in length, this probe is extremely resistant to noise sources that can severely limit other interferometers, even when using poor-coherence semiconductor laser sources.

The high sensitivity, excellent definition, and simple construction of this sensor make it a candidate for a parallel system in which a large number of spots on the sample are measured simultaneously. It is the eventual aim of this work to be able to make two-dimensional photothermal images at much greater speed than the present very slow rates.

INTRODUCTION

Photothermal measurements provide a noncontacting, nondestructive means for determining various properties of a sample such as film thickness, quality of thermal bonding, the presence of flaws, and material composition. We have also demonstrated in past work that the same techniques can be used to obtain images of regions of high rates of surface recombination in semiconductor materials. These properties are inferred from the periodic surface temperature or thermal expansion of the sample when heated by a periodic energy source, such as a modulated laser beam or an ac current. The temperature or surface displacement variation over a surface can be displayed as a grey scale image. Additionally, individual points on the surface can be interrogated for further information; for instance, since the thermal diffusion length varies inversely as the square root of the frequency, it is possible to vary the frequency and obtain information about the depth of a flaw.

It is the eventual aim of this work to obtain high-quality, high-definition photothermal and photoacoustic images of solid substrates, and to increase the speed at which such images can be taken from the present very slow rates. Typically, such imaging processes involve the use of relatively insensitive detectors, and hence long integration times are required. Furthermore, the images are normally obtained by mechanically scanning a single sensor over the area to be imaged, an inherently slow process. The slowness of the photoacoustic and photothermal imaging modes presently available is a major disadvantage, often glossed over in research papers, which seriously limits the range of applicability of these techniques to practical problems. It is
therefore imperative to develop a probe which can eventually lead to parallel processing. Only in this way can we hope to obtain images in frame times comparable to TV rates or even in a few seconds, which would be adequate. We believe that by following this approach, we have the opportunity to make photothermal and photacoustic imaging have a far wider range of practical applications than it does at present.

In the past, we developed a new type of acoustic transducer, operating in air, to detect the thermally-induced acoustic waves in the air above the sample. With these air transducers, we have measured the thickness of thin films of silicon dioxide and of ferrite-filled epoxy coatings on magnetic recording disks, surface recombination rates in semiconductors, and disbonding between silicon devices and heat sinks. These measurements provide good sensitivity to the surface temperature and an excellent ability to discriminate against sample vibration, but they suffer from the drawbacks that they have a very narrow bandwidth due to the enormous acoustic impedance mismatch between the transducer and the air, and the resolution, of the order of 300 \( \mu \text{m} \), is poor when compared to an optical probe. A related approach, which does not suffer the bandwidth limitations, is to measure the thermal gradient in the air above the sample surface by measuring the deflection of a laser beam passing over the heated area. However, this method has only one-dimensional resolution and would not be appropriate when trying to form high-resolution, two-dimensional images.

Our optical probe gets around the problems of definition and bandwidth by using a focused laser beam to measure the thermal expansion of the sample interferometrically. The probe resolution along the surface is on the order of a few micrometers, as determined by the optical lens, and it is capable of making measurements from a few hertz up to several hundred megahertz. Additionally, because it employs semiconductor lasers which are easy to modulate, and because of its simple design, it is suitable for incorporating into a parallel processing system for increasing the speed at which photothermal images can be formed. As we have already discussed, the need for long integration times in photothermal measurements severely limits their speed. In order to increase the practical applications of these techniques, a parallel process must be developed in which the integration time is between frames or between lines, at the very least.

**EXPERIMENTAL SYSTEM AND RESULTS**

The layout of our probe is shown in Fig. 1. The output from the laser passes through the polarizing beam splitter and into a birefringent Wollaston prism with its optical axis oriented at 45° to the direction of polarization. The prism splits the light into two equal-amplitude components, angularly separated by 0.5°, which are focused to two spots on the sample. The reflected beams retrace their paths and interfere on a photodetector. One of the two beams is focused onto an unheated region of the sample and serves as the reference. The other beam is focused onto the heated area where the periodic thermal expansion phase modulates it. More generally, if the two beams are closely separated, a differential signal is obtained which is proportional to the slope caused by periodic heating of the substrate. Upon interfering with the reference beam, the phase modulation is converted to an intensity modulation and the light of the photodiode is of the form

\[
P = P_0[1 + \cos(kAL) + \sin(kAL) \cdot 2kd \cos(\omega_s t + \phi)]
\]

where \( P_0 \) is the optical power in each arm, \( k \) is the optical wave number, \( AL \) is the static path length difference, and \( d \) and \( \omega_s \) are the amplitude and frequency of the thermal surface displacement.
There are a number of features of this optical probe which enable us to obtain remarkable sensitivities. First, both beams pass through the same optical components and, in fact, are nearly coincident throughout. This makes the system highly resistant to vibrations and drift. In the past, such vibrations have plagued attempts at using a Michelson interferometer to measure surface displacement. In that experiment, the reference mirror had to be mounted on a piezoelectric wafer and a feedback system had to be developed to stabilize the interferometer. In the same manner, placing the reference beam on the sample near the sensing beam eliminates the long-wavelength disturbances caused by sample vibrations.

Additionally, the probe is insensitive to phase fluctuations of the laser since both the sensing beam path and the reference beam path are equal in length. We can see from Eq. (1) that if $\Delta L$ were large, any fluctuation in emission wavelength or change in $k$ would show up as a fluctuation in the large dc current at the detector output. The Wollaston prism consists of two quartz wedges with their optical axes oriented at 90° so that the delay through the center of the cube is the same for each arm.

As an example of how serious the phase noise effects can be, Fig. 2 shows the measured noise level of our probe as a function of the path length difference. In these measurements we used a 3 mm coherence length, single-mode semiconductor diode laser and took the data at the points where $k\Delta L = (n + 1/2)\pi$. Clearly, in an unbalanced configuration, the phase fluctuations are the dominant source of noise. The theoretical curve, based on reference 4, represents the intensity noise expected from a single Gaussian mode laser of the same coherence length. The measured performance is worse because our laser actually has low-level fluctuating output over a wide range of frequencies (10-15 dBs weaker than the main signal). When, as has been done in the past in related applications, a calcite wedge was used in place of the Wollaston prism ($\Delta L = 200 \mu m$), the noise increased by 45 dBs.

It is also apparent that, when properly adjusted, this optical probe achieves shot-noise-limited sensitivity despite its use of semiconductor laser sources with poor coherence. The ability to use semiconductor lasers is an important feature of this probe. These lasers give a low-amplitude noise output enabling us to approach shot-noise-limited detection. Additionally, the ease with which these lasers are modulated provides a convenient means to extend the bandwidth of the probe above that of the photodetector or the lock-in amplifier. High-frequency photothermal effects can be measured by modulating...
Fig. 2. Total measured noise power (top trace) and theoretical noise power due to phase noise of the laser plus shot noise at the detector. Both curves are normalized by the shot noise.

the laser at a frequency close to that of the heat source and detecting at the difference frequency.

The maximum sensitivity is governed by the shot-noise power of the photodetector given by $2qP_0R_iB$, where $n$ is the detector response (mA/mW), $R_i$ is the load resistance, and $B$ is the detection bandwidth. By comparing with Eq. (1), the minimum detectable displacement is

$$\min_{d} = \sqrt{\frac{qB}{n^2P_0k^2}}$$

For $5 \text{ mW}$ of laser power and a response of $0.5 \text{ mA/mW}$, $d_{min}$ is $2.3 \times 10^{-5} \sqrt{B} \text{ m}$. In typical cases where the heating frequency is chosen so that the diffusion length is $20 \text{ m} \mu$, this might correspond to a temperature fluctuation of $10^{-5} \sqrt{B} \text{ C}$. While these values appear remarkably small, it should be pointed out that they correspond to periodic displacements rather than topographical features. In comparison with other techniques, these sensitivities are as good as or better than most noncontacting photothermal methods, but not as sensitive as transducers in direct contact with the sample.

Using this sensor, we were able to measure defects in current-carrying conductors which were heated by passing an ac current through them; the resulting temperature profile indicated the regions of high electrical resistance associated with flaws. As shown in the contour plot of Fig. 3, the temperature increase near the defect can be clearly seen. The ac temperature in this instance was around $0.01 \text{ C-rms}$. We suggest that techniques of this kind could be most important for checking the lack of adhesion of current-carrying conductors to a substrate, and for measuring metal migration effects.

In another application, we were able to measure the electrical bond quality of solder bumps connecting silicon semiconductor flip chips to current-carrying conductors on a silicon interconnect chip. It is possible, as we have shown in other work, to use a 50 MHz acoustic microscope to check for
completely disconnected bonds. It is far more difficult, however, to check for partially connected bonds, ones which are liable to develop an open circuit at a later stage. A schematic of one of the devices we have examined is shown in Fig. 4. Our basic idea is to heat the poorly-contacting solder bumps by passing an ac current through them, or by using a periodically-modulated laser source which penetrates through the silicon substrate to heat them. We determine which bumps are poor by looking for hot spots.

The small size (approximately 100 μm square) and high density of the bumps, as well as the fact that they are obscured by a thick wafer of silicon, make it extremely difficult, if not impossible, to obtain an image with standard photothermal techniques. Any method based on measurements at the exposed surface of the silicon substrate or that of the flip chips would have to contend, due to thermal diffusion, with attenuated and overlapping signals from the multiple bumps. However, by using the semiconductor diode laser as a source for a 1.3 μm wavelength probing beam, to which silicon is transparent, we were able to view the solder bumps directly. Figure 5 shows the thermal image of a row of bumps heated by passing a 20 kHz current through them. The seven bumps at the left make good electrical contact and are not visible. Of the solder joints on the right side, two are particularly poor, being about ten times higher in temperature than the others. Furthermore, information can be learned about the vertical location of the flaw by comparing images from the top and bottom surfaces and by measuring the phase lag.
Fig. 4. (a) Top view of device showing pattern of solder bumps. (b) Side view of device.

Fig. 5. Thermal image of a single row of solder bumps heated electrically, as a function of frequency, between the heat source and the point of detection.

FURTHER DEVELOPMENTS OF THE MEASURING TECHNIQUE

We are presently reconfiguring the system with a laser heating source to measure flaws and cracks in ceramics and other materials, as well as the solder bumps already discussed.
One minor difficulty, which does not give too much of a problem in practice because it requires relatively uncritical alignment, is that for optimum sensitivity, the path length difference between the two beams must correspond to a quarter wavelength. We have constructed an electro-optic phase shifter of PLZT ceramic which can introduce a periodic phase modulation between the two beams at a frequency $\Omega$. Output signals can now be obtained at frequencies $\omega \pm \Omega$; by filtering one of these signals, we can obtain an output which is independent of path length.

Finally, the simplicity of this optical probe makes it possible to incorporate it into a parallel system. This can be done by replacing the probe laser and the photodetector with a distributed light source and a detector array, each element of which images a separate point on the sample. By modulating the light source at the same frequency as the heating source, the thermal expansion signal will be converted to a small change in the DC light level. The CCD camera essentially provides 30 ms of signal averaging per frame. By reversing the phase of the light modulation after 30 ms and subtracting successive frames, the thermal effects will add while the background DC will cancel. Longer integration times can be achieved by averaging many frames.

ACKNOWLEDGMENT

This work was supported by the Department of Energy under Contract No. DOE DE-AS03-81-ER10865.

REFERENCES

AN ACOUSTIC EMISSION DIGITAL DATA ACQUISITION WORKSTATION

John A. Johnson, Basil A. Barna, & Randy T. Allemeier
Idaho National Engineering Laboratory
EG&G Idaho, Inc.
Idaho Falls, Idaho 83415 USA

ABSTRACT

A digital acoustic emission (AE) data acquisition system has been used to record the acoustic emissions generated by individual crack initiation sites during fracture. The purpose of this research is to analyze the dynamics of crack growth and to discriminate among fracture mechanisms through pattern analysis of the received signals. Preliminary experiments were performed using a laser as an AE source for calibration, and data were collected from a tensile test of an aluminum surface-cracked specimen.

During the laser calibration experiments the AE from a focused Nd-YAG laser spot, which was translated 2.0 mm across a sample in 0.1 mm steps, were recorded and used to locate the origin of the AE. The results showed the basic capability of the system to discriminate between initiation sites less than 100 μm apart.

The aluminum fracture experiment showed the capability of an enhanced system to record real events with a response time of less than 12 ms between events. The dynamics of the initiation site evolution show the progression of the crack growth across the pre-crack tip. Research is now being pursued to both validate these preliminary results and utilize the information on the dynamics of the fracture event.

INTRODUCTION

One of the primary goals of this research is to predict conditions for initiation of crack growth. The predictions will be based on models being developed by Parks and McClintock [1]. The development and verification of these models is closely tied to fracture mechanics experiments being performed at the Idaho National Engineering Laboratory (INEL). The experimental program will point the direction for development of the models and can then be used to confirm the predictions. In the development phase, finite element methods are being used by Parks and McClintock to calculate the effects of specific crack growth conditions. These calculations require detailed information about the location of the crack initiation sites and about the types of crack growth for various material properties and geometric conditions. These details will be determined using the source location and source identification methods described in this paper. The source location tells where to conduct metallographic examinations so the extent of cracking can be identified. Then, the actual conditions for the real materials around the crack border can be compared to predicted and assumed conditions in the model.

DIGITAL AE SYSTEM

The first version of the AE detection system [2,3,4] was based on an LSI 11/73 microcomputer. The system was configured to record the AE events generated by
individual crack initiation sites during fracture and had the capability to determine AE sites with a precision of better than 0.2 mm at a typical maximum rate of 15 events/second for four channels of acquisition.

A second generation system based on a MicroVAX II was designed using the experience gained with the first system. The basic data acquisition hardware remained the same but the faster, larger computer provided improved acquisition speed and larger data storage capability.

The main weaknesses in the first system were small data storage capabilities and slow speed. Data were acquired from four channels at a maximum rate of 32 megasamples per second. Each data set could consist of up to 512 values per channel or 2K values per AE event. The system was limited in speed and number of events which could be stored by the capacity of the LSI 11/73 computer. The maximum event rate of the system was about 15 per second (depending on the number of channels and number of points per event per channel) with a total maximum number of data points limited to less than 64K or 32 events at 2K values per event. A typical test run will result in hundreds of AE events, some occurring in rapid succession, and the first system can handle neither the speed nor the quantity of data. The new system can acquire at a speed of greater than 80 events per second and with a total number of data points greater than 8 million.

System Configuration

A block diagram of the AE system is shown in Figure 1. The A/D converters are 8 bit flash digitizers which can be operated at up to a 32 MHz sampling rate. The system has capability for digital storage of signals on up to four independent channels with the first channel acting as the timing reference for analysis of the data. Once armed, the system constantly digitizes the output of all active channels until the receipt of a "stop" trigger generated by the signal level in channel 1 crossing a predetermined threshold. After receipt of the stop trigger, a specified number of pre- and post-trigger samples are transferred to computer memory, and the system is cycled and armed for the next event. In addition to storing the digitized signal received by the transducer, the software also reads the system clock and records the time at the beginning of the data transfer.

To decrease the "dead time", all the data are stored in computer memory rather than on a mass storage device such as a disk. The actual dead time of the system depends on the number of channels which are active and the length of each data set. A typical dead time using an LSI 11/73 computer is 65 ms for a four channel setup collecting 161 points of data per channel. With that computer, a total of 32 Kbytes of data can be collected before filling the memory. Using a MicroVAX II computer, the dead time is reduced to 12 ms and the total amount of data is increased to approximately 8 Mbytes.

Laser Source Calibration Experiments

The capability of the system to determine source location was investigated using a laser to simulate AE events in precisely controlled locations. The first series of these experiments used two transducers from a design by Proctor [5] affixed to a steel fracture specimen 6 mm thick. A Nd-YAG laser beam was then used to generate simulated AE signals on the edge of the specimen between the two transducers. This configuration, shown schematically in Figure 2, is similar to that of the surface-cracked specimens used in the fracture mechanics experiments. Twenty-one data sets were collected with a spacing of 0.1 mm between spot positions.

Since the path length between source and transducer is a function of the source location, the signals detected in each channel will show differing time delays that depend on source location. Examples of the type of data collected are shown in Figure 3. The data were analyzed for location of the AE source by solving the equations which describe the time of arrival differences between the active channels. With the laser path on an exposed surface of the specimen for this first set of experiments, it was possible to simplify the experiment and only use two
Figure 1. System block diagram.

Figure 2. Laser generated AE for calibration.
channels of data to calculate the source positions. In the general case, it is necessary to utilize all four channels for unambiguous location in three-dimensional space.

The results are plotted in Figure 4. The best straight-line fit to the data give a slope of 0.998 ± 0.015 and an intercept of 0.004 ± 0.017. The largest error was 0.086 mm and the average error was 0.032 mm. Possible sources of the error include measurement errors in locating the transducer positions and the finite size of the laser spot. Another possible source of error is the method of determining the time shift between the slightly different waveforms received by the two transducers. The differences are thought to be due to the individual characteristics of the transducers.

Aluminum Fracture Experiment

After the calibration experiments, data were acquired with the system from aluminum surface-cracked specimen which provided actual AE for evaluation. The intent of this experiment with an actual fracture specimen was to determine how the system would respond to real events. The rapidity and pattern of real event generation, spurious electrical and mechanical noise, and the spectra of the AE all affect system performance and the parameters used to control the data collection.

The specimen had a 30 mm EDM notch from which a small fatigue crack was grown before the test. The specimen was then instrumented with four AE transducers as shown in Figure 5. During the test, the specimen was pulled in a direction perpendicular to the crack face until a predetermined load was reached. The digitized output of a typical AE event is shown in Figure 6. As in the laser source experiments, the time shift between the signals is indicative of the AE source location.

The system recorded 280 events until the maximum load was reached. An additional 72 events were recorded while the load was held; data acquisition was terminated when the load was released. The total amount of data taken was the 352 events X 4 channels/event X 40 µs/channel X 8 samples/µs or 450,560 bytes of data. Had the original LSI 11/73 system been used, only 51 events could have been recorded before filling the memory of the system.

A complete analysis of data is currently underway. An automated processing system is being designed to determine the arrival time in each channel and to calculate the location of the event. However, a preliminary look at the data indicated that only 5 of the first 34 events were significant. The others were either electrical noise or AE with nearly identical arrival times which was probably due to the specimen grips. Of the remaining events, about one out of every twenty was electrical noise and the rest appeared to be real AE events from the area of the crack. All of the very large events were analyzed and the apparent source location determined assuming that the event was in the plane of the crack. A nonlinear least-squares fitting program determined the position along the crack and the depth
of the event. The three events with the smallest error in their locations, as determined by the fitting program, were used to pick several areas of the specimen for destructive evaluation.

SUMMARY

A digital AE data acquisition system was used to collect preliminary data from both laser calibration experiments and an aluminum fracture specimen. The calibration experiments show a resolution between adjacent sources of better than 100 μm. The system also showed apparent sensitivity to actual brittle fractures in an aluminum sample in that the locations of the calculated fracture sites are consistent with the known pre-crack extent before significant crack growth. System response times with the configuration used in these experiments are on the order of 12 ms.

Further experiments are now required to verify these preliminary results. An error analysis is needed to quantify the sensitivity of the location calculation to errors in the measured transducer positions and the effects of the number of channels used as input to the calculation. Once the sources of error are further identified and quantified, the source location information which the system provides can be correlated with a time history of the free surface strain field surrounding the crack observed using moire techniques. These correlations will provide insight into the dynamics of the fracture event. Future work will also attempt to extend the analysis capability beyond source location to characterizing the type of fracture generating the AE. Other research [6] indicates it should be possible to discriminate among mechanisms such as inclusion fracture and ductile tearing from an analysis of the received waveforms. Plans for an automated pattern-recognition system for the AE waveforms are being developed.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Energy Research, Office of Basic Energy Sciences, under DOE Contract No. DE-AC07-76ID01570.
Figure 5. Fracture test setup showing the locations of the transducers and pre-crack.

REFERENCES


Figure 6. Digitized output of a typical fracture event during the aluminum surface crack tensile tests.


CRACK CHARACTERIZATION BY SCATTERING METHODS

J. D. Achenbach
The Technological Institute
Northwestern University
Evanston, IL. 60201, U.S.A.

ABSTRACT

Geometrical irregularities and material inhomogeneities in the vicinity of a crack can significantly affect the scattered fields that are generated by incident ultrasonic waves. In this paper we report some results that display the effect of deviations from a mathematically perfect simple crack geometry. Most of the results have been obtained by formulating a set of singular integral equations for the fields on the boundaries of the scattering obstacles. These equations have been solved numerically by the boundary element method, and the scattered fields have subsequently been obtained by using representation integrals.

INTRODUCTION

Flaws in solid bodies can, in principle, be detected and characterized by their effect on an incident pulse of ultrasonic wave motion. In actual testing situations there are, of course, several complicating factors which may make it impossible to produce a signal that provides enough information to completely characterize the flaw.

Analytical and numerical results for ultrasonic wave scattering by a crack provide valuable quantitative information for methods to detect and characterize cracks. From the mathematical point of view an open crack is a surface which does not transmit tractions. The implementation of this statement in a mathematical formulation generally involves a number of idealizations. Thus, the faces of a perfect mathematical crack are taken as smooth. It is also assumed that they are infinitesimally close, even though it is supposed that the faces do not interact with each other.

Real cracks, particularly fatigue and stress corrosion cracks have rough faces, which may contact each other. Sometimes there is not a single crack, but rather a configuration of a principal crack and an adjoining satellite crack, for example, a macrocrack and a neighboring microcrack. It also frequently happens that other smaller scatterers are located near a crack. The secondary scattering from these inhomogeneities will also affect the overall scattered field. In recent work reported in this paper we have investigated to what extent these complicating features of real cracks or systems of cracks can be accounted for in an analysis.

There are few exact closed-form analytical solutions to elastodynamic scattering problems. In a rigorous approach, often the best that can be done is to reduce the mathematical formulation to a form which is suitable for numerical work. Closed-form solutions can be obtained at high frequencies (Kirchhoff approximation or geometrical diffraction theory) or at low frequencies (Rayleigh approximation). However these approximations are of limited value for scattering in the mid-frequency range in which characteristic wavelengths of an incident displacement pulse are of the same order of magnitude as a characteristic length of the flaw. In that range it is necessary to use numerical methods; for example, numerical schemes to solve systems of governing integral equations, finite difference techniques,
finite element techniques, T-matrix methods, etc.

As shown in Ref.[1], the effect of a crack on an incident field of ultrasonic wave motion can be nicely displayed in full-field snapshots, with the aid of time-domain finite difference calculations. Some results are shown in this paper. For crack characterization, only the fields at a few well selected points of observation are, however, generally required. These are generally more efficiently calculated by the use of the boundary element method. This method can also be used when the crack is not perfect. Results for a configuration of a crack with a near-tip circular void on a near-tip circular inclusion are also discussed in this paper.

TIME-DOMAIN DIFFERENCE CALCULATIONS

The finite difference method used is explicit in time. The details have been presented in Ref.[1]. In Ref.[1], use is made of previously unreported difference formulations for some of the special nodes contained in the numerical domain. A radiation condition which allows free transmission of energy from the numerical domain for body waves as well as surface waves is successfully employed.

![Fig. 1](image)

**Fig. 1.** (a) Configuration of plate with surface-breaking crack; (b) shape of incident pulse

The configuration is shown in Fig. 1. A transducer on the upper face of the slab generates a transverse wave which propagates towards the surface-breaking crack. All fields are taken as independent of z, which reduces the problems to ones of two-dimensional plane-strain scattering.

The transducer has been simulated by choosing the surface tractions in such a way that they produce a beam of transverse wave motion, of finite width and finite duration, which traverses the slab at a 45° angle, and is directed towards the mouth of the surface-breaking crack. The transverse displacement pulse illustrated in Fig. 1(b) was used. The tractions applied at the transducer/slab interface correspond to those for a pulse in an unbounded medium of the form shown in Fig. 1(b), which passes the interface "window" and produces tractions along the interface. These are just the values used in the numerical simulation. Since the plane wave passes the interface at a 45° angle no shear tractions are induced. A time delay is incorporated into the incident pulse so that at the onset of the numerical iterations, displacements and particle velocities may be set to zero.

To complete the formulation of the problem, conditions at the ends of the slabs must be specified. As the slabs are semi-infinite in extent, the displacement waves are outgoing. Hence, in the truncated numerical domain a radiation condition is employed.
Fig. 2. Snapshots for homogeneous plate with a crack: (a) $t = 13.5$ μs; (b) $t = 20.25$ μs; (c) $t = 27$ μs; (d) $t = 40.5$ μs.

Fig. 3. Time histories of normal velocity at the midpoint of the transducer: (a) no crack, --- crack present.
The numerical results are displayed in two ways, (1) spatial displacement distributions (snapshots) at a specified time, and (2) time histories of the normal component of the particle velocity at the midpoint of the transducer/plate interface. A spatial displacement distribution is obtained by depicting displacements by vectors emanating from the corresponding nodal points, but only at nodes where the displacement magnitude is greater than a specified value. This yields a snapshot of the displacement distribution generated by scattering of the incident pulse, both in magnitude and direction. The normal component of the particle velocity at the midpoint of the transducer/plate interface has been chosen for the time-history display.

Figure 2 shows snapshots for the case of a homogeneous plate with a crack. The four pictures represent the displacement field at times 13.5 µs, 20.25 µs, 27 µs, and 40.5 µs. Only displacements with a magnitude greater than 0.054 (scaled with respect to the maximum amplitude of the incident wave) are shown, while displacements having a magnitude greater than 0.18 are represented by vectors with an arrow in the direction of the displacement.

The values of Δt and h are Δt = 0.09 (scaled with respect to t) and h = 0.05 (scaled with respect to the crack length). The calculations were carried out for r = 1 µs, a = 1.27 cm, H = 2.54 cm, c_L = 6 x 10^5 cm/s, and c_T = 3 x 10^5 cm/s. The midpoint of the transducer/plate interface is defined by x_M = (-2,2). The right and left artificial boundaries are positioned at x_BR = 2.5 and x_BL = -6.5, respectively.

The transducer/plate interface is represented by the line parallel to the upper surface of the plate. The incident displacement field is illustrated in Fig. 2(a), in which the transverse pulse, the Rayleigh surface wave, and the cylindrically spreading displacements caused by the edges of the transducer can be seen. The surface breaking crack at the bottom surface causes a reflection of the 45° shear pulse while leaving the surface wave on the upper plate relatively unchanged. The final snapshot shows that the radiation conditions applied at the artificial boundaries are effective. It is noted that the Rayleigh surface wave is also traveling out of the domain on the right side. Figure 2(d) also shows the reflection of the transverse pulse from the upper plate surface and the surface wave on the lower surface, which is reflected from the crack.

Figure 3(a) compares the time histories of the normal velocity at the midpoint of the transducer/plate interface, for the cases with and without a crack. The most obvious differences are the sharp peaks centered around 30 µs for the plate with a crack. This pulse represents the reflection of the transverse pulse by the crack faces. The velocity of the incident pulse caused by the tractions applied at the transducer interface is represented by the peaks centered around 7.5 µs. A small peak near 25 µs can be identified as the diffraction from the crack tip.

REFLECTION AND TRANSMISSION BY A FLAW PLANE

In Ref.[2], the interaction of elastic waves with a planar array of periodically spaced cracks of equal length was investigated. Normal incidence of both longitudinal and transverse waves was considered. The results of Ref.[2] can be brought into perspective with investigations for the reflection of elastic waves by a flaw plane, when the flaw plane is viewed as a region where separation zones and contact zones can both exist, and friction may be significant over the contact zones.

It is not difficult to show that a periodic array of identical compact inhomogeneities whose geometrical centers are located in a plane interior to an elastic solid, acts as a homogeneous plane of reflection and transmission of incident waves, at least at low frequencies. A typical example is provided by incidence of a plane elastic wave on a single layer of periodically spaced spherical cavities. Reflection and transmission coefficients for that configuration have been obtained by Achenbach and Kitahara [3]. Their results show that for an arbitrary angle of incidence, an incident plane wave gives rise to an infinite number of reflected and transmitted longitudinal and transverse wave modes. The higher-order
modes have cut-off frequencies below which these modes are evanescent. Below the
first cut-off frequency only the zeroth-order modes are propagating, and these modes
are reflected and transmitted homogeneous plane waves.

The analysis of Achenbach and Kitahara [3] exploits the periodicity of the
array of spherical cavities. The reflection and transmission problem was
formulated rigorously. Reflection and transmission coefficients were computed by a
novel application of the Betti-Rayleigh reciprocal theorem. They are expressed as
integrals over the surface of a single inhomogeneity, in terms of the field
variables on that surface. The fields on the surface of that inhomogeneity are
governed by a system of boundary integral equations which have been derived in some
detail. This system was solved numerically by the use of the boundary element
method.

**DISPERSION RELATION**

It was shown by Achenbach and Kitahara [4] that the results for reflection and
transmission by a single layer of inhomogeneities can be used to obtain a dispersion
relation for wave propagation through multiple arrays of inclusions. The distance
between inclusions in the $x_3$-direction is denoted by $h$. The centers of the
inclusions are now located at positions

$$x_1 = ma, x_2 = nb \text{ and } x_3 = \ell h \quad (m,n,\ell = 0, \pm 1, \pm 2, \cdots).$$

The material properties of all the inclusions are assumed to be the same. The wave motion is restricted to the
frequency range of the zeroth order modes.

On the basis of Floquet theory or the one-dimensional version of Bloch theory,
we obtain a dispersion equation in the form

$$\cos(qh) = \left[\frac{(R^L)^2 - (T^L)^2}{2T^L}\right] + 2\cos(k_L h)/2T^L.$$

Here $R^L$ and $T^L$ are the reflection and transmission coefficients. For a given
frequency, i.e., for a given value of $k_L$, the Floquet wavenumber $q$, which may
become a complex number, can be obtained from eq. (1).

**SCATTERING BY MULTIPLE CRACK CONFIGURATIONS**

The perfect mathematical crack model is acceptable for a real crack, provided
that the latter's faces are slightly separated and that the length characterizing
the crack-face roughness is much smaller than the dominant wavelength of an incident
wave pulse of ultrasonic wave motion. When the faces of a crack are in contact and the
crack is actually partially closed, or when there are two neighboring cracks, the
model must be adjusted. In this paper we assume that a partially closed crack can
be represented by a configuration of two neighboring cracks. We consider the
configuration which is depicted in the insert of Fig. 4.

The mathematical formulation is two-dimensional, and for a state of plane
strain. The details can be found in a paper by Zhang and Achenbach [5]. The usual
displacement boundary integral equation formulation for scattering by volume
scatterers, disintegrates when the scatterer is reduced to a crack. Hence a
traction boundary integral equation is used in this paper. The mathematical
formulation for a single crack is well known. Here it is extended to a two-crack
configuration, the result being a system of four coupled singular integral equations
for the general case. These boundary integral equations are highly singular, and
they cannot be solved directly by numerical methods. To circumvent this difficulty
a special approach has been developed by which the higher order singularity has been
reduced to a Cauchy-type singularity. Numerical results for scattered fields for
several configurations have been obtained in Ref. [5] by the boundary element method.
The scattered field results apply to a partially closed crack as well as to two
separate but neighboring cracks.
For crack problems the scattered displacement $u^{sc}_\gamma$ can be expressed by the following representation theorem

$$u^{sc}_\gamma(x_p) = \int_{\Gamma} \sigma^G_{\alpha\beta\gamma}(x;x_p) \Delta u_\alpha(x)n_\beta ds, \quad x_p \in \Gamma.$$  \hspace{1cm} (2)

Here $\Gamma$ represents the insonified sides of the cracks. Also $x_p$ is the position vector of the observation point, $\sigma^G_{\alpha\beta\gamma}(x;x_p)$ is the Green's function, $\Delta u_\alpha$ is the displacement jump (crack opening displacement) across the cracks and $n_\beta$ is the unit normal vector of $\Gamma$. In addition, the scattered field must fulfill the radiation condition.

Substitution of equation (2) in Hooke's law and use of the relations between surface tractions and the stress components

$$f^{sc}_\alpha(x_p) = -n_\beta(x_p) \int_{\Gamma} \left( \lambda \delta_{\alpha\beta\gamma} + \mu \delta_{\alpha\gamma} \right) \Delta u_\alpha(x)n_\beta ds + \mu \sigma^G_{\delta\epsilon\beta\alpha}(x;x_p) \Delta u_\epsilon(x) n_\alpha ds,$$  \hspace{1cm} (3)

yields the representation formula for the traction components at $x_p$:

$$f^{sc}_\alpha(x_p) = -n_\beta(x_p) \int_{\Gamma} \left( \lambda \delta_{\alpha\beta\gamma} + \mu \delta_{\alpha\gamma} \right) \Delta u_\alpha(x)n_\beta ds + \mu \sigma^G_{\delta\epsilon\beta\alpha}(x;x_p) \Delta u_\epsilon(x) n_\alpha ds + \mu \sigma^G_{\delta\epsilon\beta\alpha}(x;x_p) \Delta u_\epsilon(x) n_\alpha ds,$$  \hspace{1cm} (4)

We note here that the equation (4) becomes improper if the observation point $x_p$ and the boundary point $x$ coincide. In this case the kernel function of (4) would be highly singular and non-integrable. Hence we cannot take the limit $x \to \Gamma$ at this point. In order to avoid this difficulty we first divide $\Gamma$ into $N$ elements and we take the unknown quantity $\Delta u_\alpha$ in (4) as constant over each element. Then equation (4) can be written in the following discretized form:

$$f^{sc}_\alpha(x_p) = -n_\beta(x_p) \sum_{j=1}^{s+1} \Delta u_\delta(x_j) \int_{s_j}^{s_{j+1}} \left( \lambda \delta_{\alpha\beta\gamma} + \mu \delta_{\alpha\gamma} \right) \Delta u_\alpha(x)n_\beta ds + \mu \sigma^G_{\delta\epsilon\beta\alpha}(x;x_p) \Delta u_\epsilon(x) n_\alpha ds,$$  \hspace{1cm} (5)

where $s_j$ denotes the first node and $s_{j+1}$ denotes the second node of the $j$-th element, while $x_j$ is its mid-point.
Equation (5) forms the basis for a numerical method which has been discussed in considerable detail in Ref. [5]. Some results are shown in Fig. 4, where the backscattered displacement field is plotted versus d/a. Here b/a = 0.3 and k_a = 1. It is not surprising that the backscattered field decreases with increasing d/a even though the changes are small. Also shown is the corresponding curve for a single crack of half-length \( \bar{a} = a + b + d/2 \). The comparison of the two curves shows clearly that a partially closed crack, with relative closure length d/a, may be expected to generate a smaller backscattered field.

![Diagram](image)

**Fig. 5.** A crack with a void, and a crack with an inclusion.

**THE EFFECT OF A VOID OR AN INCLUSION**

Scattering by a configuration consisting of a crack and a void (or an inclusion) can also be studied by reducing the problem formulation to a set of singular integral equations, and by subsequently solving these equations by the use of the boundary element method. Here we present some preliminary results for the configurations shown in Fig. 5. For grazing incidence of a transverse wave, the backscattered fields in the \( x_2 \)-direction are shown in Fig. 6. It is noted that there is a significant difference between the effects of a circular inclusion and a circular void.

**CONCLUDING COMMENTS**

This paper has reported on work that is in progress on the effects of deviations from perfect crack geometries on the scattered ultrasonic fields. The best method for elastodynamic scattering problems (for compact inhomogeneities or clusters of inhomogeneities) uses Green's functions techniques to formulate a system of singular integral equations for the fields on the bounding surfaces of the inhomogeneities. The equations can be solved by using the boundary element method. Representation integrals can then be used for the scattered field. These calculations are generally carried out in the frequency domain, and the Fast Fourier...
Fig. 6. Back scattered fields for the configurations shown in Fig. 5.

Transform is subsequently used to obtain time domain results. It is, however, also possible to follow essentially the same steps, and formulate and solve the problem directly in the time domain. Further work is in progress to determine the range of geometrical parameters and wavelengths, which would give rise to serious disfiguration of the component of the scattered field produced by the crack.

ACKNOWLEDGMENT

This paper was written in the course of research sponsored under Contract DE FG02-86-ER13684 with the Department of Energy, Office of Basic Energy Sciences, Engineering Research Program.

REFERENCES

5. C. Zhang and J. D. Achenbach, "Scattering by Multiple Crack Configurations," submitted for publication.
A basic concept of in-process control of residual stresses and distortion in weldments is presented. Significant reductions of residual stresses and distortion can be achieved by taking proper actions during welding while plastic strains are being formed. As an example of in-process control, experimental and analytical studies were made of reducing the opening of the root gap during gas metal arc welding of butt joints. The objective of the research was to reduce forces acting on tack welds so that a number of tack welds that act as major disturbances can be reduced or even eliminated. It has been found that additional heating by a set of oxyacetylene torches alongside the welding arc can produce thermal stresses that reduce the root gap opening. The optimum heating condition is to heat a wide region at relatively low temperatures.

1. BACKGROUND AND BASIC CONCEPT

A major problem associated with arc welding is that related to distortion and residual stresses. Because a weldment is locally heated by the arc, complex thermal stresses occur during welding, and residual stresses and distortion remain after welding is completed. These stresses and distortion cause complex consequences, most of which are detrimental, during fabrication and service. They include the formation of joint mismatch during welding, cracking, and premature failures of welded structures. There is always a strong interest in the industry for developing efficient means for controlling and reducing these stresses and distortion.

Subjects related to residual stresses and distortion in weldments have been studied by many investigators since the 1930's. Because residual stresses and distortion are produced by complex mechanisms, studies before around 1960 were based on experiments and/or analyses of very simple cases. Since then computers have become widely available, and a number of investigators have advanced the technology of analyzing residual stresses and distortion by use of new computational techniques such as the finite method. Since the late 1940's, Masubuchi has been very active in advancing the knowledge of residual stresses and distortion in weldments [1]. However, until recently no fundamental and systematic research was performed with an objective of developing new technologies of aggressively reducing and controlling them. The current DOE funded research program entitled "In-Process Control of Residual Stresses and Distortion" is unique in that researchers for the first time have tried to advance the state-of-the-art from just analyzing residual stresses and distortion to controlling and reducing them.
1.2. In-Process Control of Residual Stresses and Distortion

The reason why real-time control is important for reducing residual stresses and distortion can be understood by studying mechanisms of their formation.

Figure 1 shows schematically how a rectangular plate deforms when arc welding is performed along its upper longitudinal edge. Since temperatures are higher in areas near the upper edge, these areas expand more than areas near the lower edge causing upward movement of the center of the plate, $\delta$, as shown by Curve OA. The most important stress component is the longitudinal stress, $\sigma_x$. Stresses in regions near the weld are compressive, because thermal expansions in these areas are restrained by the surrounding metals at lower temperatures. Since the temperatures of the regions near the weld are quite high and the yield stresses of the material are low, compressive plastic strains are produced in these regions.

When welding is completed and the plate starts to cool, it deforms in the opposite direction. If the material was completely elastic during the entire period of the heating and cooling cycle, the plate would deform as shown by Curve OAB'C' returning to its original shape with no residual distortion. However, this does not happen during welding of a real material, be it steel, aluminum, or titanium. As a result of the compressive plastic strains that are produced in areas near the upper edge, the plate continues to deform after passing its original shape, resulting in the negative final distortion, $\delta_f$, when the plate cools down to its initial temperature (see Curve OABC).

The most effective way of reducing distortion is to control the formation of plastic strains near the weld. The difficulty here is that the necessary control must be made during welding. If the control performed is correct, the final distortion will be reduced. If the control performed is incorrect, on the other hand, the final distortion will be increased. In order to perform correct controls consistently, the following must be done:

1. Proper Prediction. One must have proper knowledge about (a) how the weldment being studied deforms and (b) how to perform control to change the distortion.
2. Proper Sensing and Control. One must also have a proper device or devices for sensing if what should happen is actually happening. If not, he/she must make the necessary changes.

1.3. Objective of the Current Research Program

The objective of the current research program which started in January 1985 is to perform fundamental research for developing technologies of in-process control of residual stresses and distortion in weldments. The work performed thus far covers the following tasks:

Task 1: In-process control of residual stresses and distortion of girth-welded
pipes;

Task 2: Development of technologies for minimizing and eliminating, if possible, tack welds in butt joints.

Task 2 consists of the following subtasks:
Subtask 2-1: Analytical and experimental studies on metal movements during welding and forces acting on tack welds;
Subtask 2-2: Research on methods of minimizing interference from tack welds.

The work has been performed as a part of the collaborative research between M.I.T. and INEL (Idaho National Engineering Laboratory) on welding automation. This paper covers the work performed under Subtask 2-1. Results of Task 1 are in Reference [2] and [3], while results of Subtask 2-2 are in Reference [4].

When a butt joint is being welded, unwelded portions of the joint move causing a joint mismatch. A common method of coping with this problem is to have tack welds. This can be done relatively easily in welding fabrication using human welders. When automatic welding machines or robots are introduced, tack welds still have to be made by a human welder adding an extra cost. During the root-pass welding, tack welds act as major disturbances, and they frequently cause lack of fusion, slag inclusion, and other types of defects. It would be very useful if techniques are developed for reducing the joint mismatch so that the number of tack welds can be reduced or even eliminated completely.

2. CONTROL OF JOINT MISMATCH DURING BUTT WELDING

2.1 Modeling of Joint Mismatch

When a long butt joint is welded from one end to the other, unwelded portions of the joint move due to thermal stresses causing joint mismatch or opening or closing of the root gap. The root gap normally increases during gas metal arc welding of plates in low-carbon steels and aluminum alloys.

In order to analyze mechanisms of joint mismatch, a simplified model as shown in Figure 2 has been developed. It is assumed that the joint mismatch is formed as a combined result of the following two phenomena:
(1) Unwelded portions of the joint ahead of the arc move due to the fact that only one edge of the plate is being welded;
(2) When the weld metal solidifies, it will prevent the reverse movement of the joint when it cools down. This occurs in a small region (of the length \( \ell_m \)) just behind the arc.

\[
\Delta \text{gap}(x) = \Theta(x) \ell_m \\
\Theta(x) = \int (l/R) \, dx \\
\text{if } (l/R) \text{ const} \\
\Delta \text{gap}(x) = (l/R) \ell_m x
\]

Figure 2. Root Gap Changes During Welding

By assuming a quasi-stationary state for the temperature field and no change of curvature after the weld metal has solidified, the change of root gap or the joint...
mismatch can be expressed as shown in Figure 2. Here $\Delta$gap $(x)$ means the change of root gap at the welding arc position, $x$, and $l_m$ means the length of the weld metal that has not yet solidified. Also, $(1/R)$ represents the curvature of the plate when it is thermally expanded.

2.2. Experimental Set-up

Figure 3 shows a typical specimen used in experimental studies. To hold two plates to be welded together, semicircular rings were welded to both ends of the plates. Strain gages were mounted on these rings so that forces acting on the rings could be monitored during and after welding. These rings simulate tack welds. When the research is completed and results of the research are used in practice, these instrumented rings may be used as devices for sensing and controlling the joint mismatch. In such a case, the rings may be mechanically clamped to the plates rather than welded to them.

![Figure 3. Experimental Set-up](image)

Figure 3 also shows a set of oxyacetylene torches used to reduce the root-gap opening caused by welding.

Conditions for welding and side heating are as follows:
- Welding current: 260 Amperes
- Arc voltage: 28 Volts
- Wire feed speed: 350 in./min
- Travel speed: 0.217 in./sec
- Tip orifice size: 0.05 in.
- Acetylene consumption: 10 cf/h

2.3. Experimental and Analytical Results

To determine the optimum side heating conditions, a one-dimensional analysis considering only the longitudinal stress, $\sigma_x$, was performed. Regarding the energy distribution used for analyzing heat flow in the plate, a point source was used for the welding arc, while a Gaussian distributed heat source was used for the oxyacetylene torch. A thermo-elastic-plastic material model was used to calculate stresses and strains in the plate. A method of successive approximation was used to determine plastic strains.
Figure 4 shows a relationship between the side heat input and the distorted curvature, \(1/R\), at the welding arc position of the plate. If the curvature is zero, the plate experiences no change of root gap during welding. As the side heating input increases, the counter balance action by the side heating input increases almost linearly, and the mode of joint mismatch changes from opening to closing at 180 cal/sec for this specific case.

![Figure 4. Curvature vs. Side Heat Input](image)

How the side heating is distributed also affects the amount of the counter balance action. Figure 5 shows that a less concentrated heat is more effective for exerting more counter balancing action. Here, the concentration coefficient, \(k\), is the coefficient which shows the concentration of heat input for a Gaussian distributed heat source:

\[
q(r) = q_{\text{max}} \exp[-kr^2]
\]

where, \(q(r)\) is the heat intensity at the distance \(r\) from the center of the heat source. A higher value of \(k\) means more concentrated heat source. When the heat is concentrated at a point, \(k\) becomes infinity.

![Figure 5. Curvature vs. Concentration Coefficient of Side Heating](image)
Figure 6 is a graph showing residual stresses caused by both welding and side heating. If the side heating is concentrated, it causes residual stresses; therefore, the plate has two peak residual stresses from welding and side heating. This is, of course, undesirable. If the side heating is less concentrated, however, it causes no additional residual stresses.

Figure 6. Calculated Distributions of Residual Stresses

These results show that the most appropriate way to reduce forces acting on tack welds is to provide side heating with low heat input and in a wide region so that a significant amount of counter balancing action can be exerted without producing additional residual stresses.

To prove the effectiveness of side heating, experiments were conducted with and without side heating. Strain values were acquired and stored in a computer by a data acquisition system. Strain data obtained on rings attached to the both ends were converted to equivalent forces. Figure 7(a) shows changes of forces acting on the rings during and after welding of a specimen without side heating. A considerable difference exists between the forces acting on the ring placed at the starting end and those acting on the ring at the finishing end. At the starting point, shrinkage took place immediately after welding commenced. At the finishing point, on the other hand, the root gap increased as welding progressed, and shrinkage took place after welding was completed.

Figure 7(b) shows similar results when side heating was employed. It is clear that the amount of root opening during welding decreased significantly by the side heating. The dotted line in Figure 7(b) is an ideal case in which no joint mismatch occurs.

Results shown in Figures 7(a) and (b) illustrate that the metal movement at the starting point is little affected by the side heating. This indicates that the measurement at the starting point can be used as a control. In other words, the joint mismatch can be minimized as long as the measurements obtained on the finishing point are similar to those obtained on the starting point.
3. CONCLUSIONS

(a) In-process actions are essential for controlling and reducing distortions and residual stresses in weldments.

(b) Side heating by oxyacetylene torches is a promising method for reducing forces acting on tack welds.

(c) A broader heating at relatively low temperatures is an effective method for exerting counter actions to reduce the mismatch caused by welding.

4. REFERENCES


SENSING AND MODELLING OF GAS METAL ARC WELDING

H. B. Smartt, J. A. Johnson, R. T. Allemaier, N. M. Carlson, C. J. Einerson, A. D. Watkins

Idaho National Engineering Laboratory
EG&G Idaho, Inc.
P. O. Box 1625
Idaho Falls, Idaho 83415

ABSTRACT

Welding automation is normally concerned with directly controlling such factors as current, arc voltage, travel speed, and filler wire speed. Several other levels of control can be identified beyond present technology including control of weld heat input, cooling rate, and bead geometry (bead width, penetration, reinforcements, and defects). This work describes the development of a steady-state model of the gas metal arc (GMA) welding process which allows independent control of the heat and mass input to the base metal. An electro-optical technique for measurement of the welding pool and groove geometry is also described along with an ultrasonic method for sensing the weld bead.

INTRODUCTION

In simplest terms, the function of any fusion welding process is the input of heat and mass to the joint to be welded. The response of the material(s) being joined to these inputs (and any post-weld treatments) determines the state of the weld and thus its properties, which determine its suitability for the desired application. The objective of a rational arc welding procedure, then, should be to provide the appropriate heat and mass input to the weld which will result in the desired properties. Welding procedures are generally based on the allowable ranges of the critical parameters, and automation, if used at all, normally deals with direct control of factors such as current, arc voltage, travel speed, and metal deposition rate.

In recent years, there has been a significant growth in the use of mechanized welding systems. The most popular alternative process is gas metal arc (GMA) welding and its variants, especially for arc welding applications using robots. The GMA process as used today, however, is subject to the same fundamental limitations as virtually all arc welding processes. Control of the process is limited to those factors to which machine builders are accustomed, such as electrode wire speed, welding speed, current, and voltage. The factors that the welding engineer would like to control, such as reinforcement area, weld heat input, microstructure, and mechanical and physical properties of the weld, are not directly controllable except in isolated cases.

This work, in collaboration with related work at MIT, is an effort to develop the fundamental basis to independently control desired physical properties of the
weld. To achieve this goal, sensing and modelling of the physical processes occurring in GMA welding are required. Toward this end, three integrated research efforts are being conducted at the Idaho National Engineering Laboratory (INEL): 1) developing a real-time process control model, 2) developing and implementing a weld vision system, and 3) ultrasonic sensing of the weld bead.

GMA PROCESS MODEL

The model developed in this work is intended for use in welding thick section carbon or alloy steel in the spray transfer mode. The requirements of the model have been set somewhat arbitrarily at controlling the weld bead reinforcement area to within ±5%, and controlling the weldment heat input per unit length of weld to ±10%. These ranges are believed to be realistically achievable and usable. At this point, the model does not contain information regarding the formation, detachment, or transfer of the droplets from the electrode wire to the weld pool.

A set of six equations provides the basis for the model:

1. \( H = \frac{EIn^*}{R} \)  
   Heat input to the base metal
2. \( G = \frac{A_wS}{R} \)  
   Reinforcement
3. \( E = E_o + nI \)  
   Power supply characterization
4. \( IE = IV_w + IV_a \)  
   Power balance
5. \( IV_w + n'IV = A_w6SHm \)  
   Power required to melt the wire
6. \( IV_a = I^2R_a \)  
   Ohm's Law

The terms in the above and following equations are listed in Table 1. These equations may be solved to give:

\[
S = \frac{E_o n^* 4 G I + n n^* 4 G I^2}{n d^2 H} \tag{1}
\]

Table 1. Definitions of Welding Terms

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H )</td>
<td>heat input per unit length (J/mm)</td>
</tr>
<tr>
<td>( E )</td>
<td>arc voltage - contact tip to workpiece (V)</td>
</tr>
<tr>
<td>( I )</td>
<td>current (A)</td>
</tr>
<tr>
<td>( \eta^* )</td>
<td>overall heat transfer efficiency</td>
</tr>
<tr>
<td>( R )</td>
<td>welding speed (mm/s)</td>
</tr>
<tr>
<td>( S )</td>
<td>wire feed speed (mm/s)</td>
</tr>
<tr>
<td>( A_w )</td>
<td>wire diameter (mm)</td>
</tr>
<tr>
<td>( E_o )</td>
<td>open circuit voltage (V)</td>
</tr>
<tr>
<td>( n )</td>
<td>volt/amp slope of power supply (V/A) (-0.0307 V/A)</td>
</tr>
<tr>
<td>( V_w )</td>
<td>voltage drop across wire length (V)</td>
</tr>
<tr>
<td>( V_a )</td>
<td>voltage drop across length of welding arc (V)</td>
</tr>
<tr>
<td>( \delta )</td>
<td>average wire density (gm/mm^3)</td>
</tr>
<tr>
<td>( \eta' )</td>
<td>fraction of total power transferred from anode spot to solid wire</td>
</tr>
<tr>
<td>( H_m )</td>
<td>heat required to melt wire (J/gm)</td>
</tr>
<tr>
<td>CTBMD</td>
<td>contact tip to base metal distance (mm)</td>
</tr>
</tbody>
</table>
and
\[ R = \frac{S \pi d^2}{4G} \]  \hspace{1cm} (2)

The current \( I \) is given by:
\[ I = \frac{E_0 \pm \sqrt{E_0^2 + 4 \left[ n - (1 - n') R_a \right] S m d^2}}{2 (1 - n') R_a - n} \frac{\left( \int_T C_p dT + h_f \right)^m}{} \]  \hspace{1cm} (3)

Two terms in equation (3) proved to be difficult to evaluate: \( R_a \), the resistance of the arc, and \( n' \), the efficiency of heat transfer from the arc to the electrode wire. In addition, equations (1)-(3) must be solved iteratively, which proved to be very time consuming for controlling the welding hardware with the computer used. As an alternative method, equation (3) was replaced with an empirical relationship for \( I \), which was derived using a statistical program, SAS. This program calculates a best-fit curve for a set of data:
\[ I = C_0 + C_1 \cdot CT + C_2 \cdot E_0 + C_3 \cdot I + C_4 \cdot CT \cdot E_0 \]
where the \( C \) terms are derived constants.

A feedback control scheme was developed to control heat input and reinforcement area by measuring the current. Referring to Figure 1, the desired reinforcement area and heat input are used as input to the model, along with the other welding parameters. The actual welding current is measured and compared to the value predicted by the model, and the difference is used to change the model. The welding speed is simultaneously adjusted to maintain the correct reinforcement area and heat input. In the block diagram (Figure 1), only the top feedback loop is presently implemented.

Welding trials in this work using the model to control reinforcement area and heat input were made using 0.89 mm diameter type E70S-6 wire, with 98% Ar-2% O\(_2\) shielding gas, on 12.7 mm thick type A-36 steel plate in a bead-on-plate configuration. The contact-tip to work piece distance was maintained constant at 15.9 mm and the power supply open circuit voltage was 32 V. All welding involved spray transfer of the molten electrode wire material to the weld pool.

The accuracy of the system has been demonstrated by measuring reinforcement areas and calculated heat inputs obtained during welding for various machine settings\(^1\). The reinforcement areas are within ±5% of the desired values, the heat inputs are well within ±1% of the desired values. The reinforcement areas were measured from macrophotographs of transverse weld sections using an image digitizing program. The heat inputs were calculated from current, voltage, and travel speed values by the control computer during welding, and assuming a 75% heat transfer efficiency.

To bridge the gap between the control of heat and mass input to the weld and sensing of such factors as weld bead cooling rate and geometry, it is necessary to develop suitable strategies for selecting the desired weldment heat and mass input. These strategies will be based to a great extent on the development of process maps as shown in Figure 2. In Figure 2, the number of weld passes required to fill a certain (arbitrary) weld joint configuration, and the material heat input limits (for HY-130 material) have been superimposed over the spray transfer region for a given voltage and initial contact-tip to workpiece distance. Extending this information over a sufficient range in four dimensions (heat input, mass input, voltage, contact-tip-to workpiece distance) results in a description of the allowed process operating space. Adding pulsed current capability (high current level and time, low current level and time) adds additional complexity.
Fracture toughness depends on weld bead cooling rate. The feasibility of controlling weld bead cooling rate in real time, independent of the reinforcement area, was examined using a set of welds made at heat inputs from 1050 to 1500 J/mm using reinforcement areas in the range of 30 to 40 mm$^2$. The weld bead cooling rates were measured by plunging thermocouples into the weld pool at the top center of the weld bead. Although the range of values included in this study is limited, it is clear that the weld bead cooling rate is related to the heat input. Future work on this problem will involve real time measurement of the weld bead cooling rate with an infrared camera in a manner similar to Lukens and Morris, and direct closed-loop control of the cooling rate with the error in cooling rate being used to signify the required change in weld bead heat input. Implementation of this scheme is shown as the second feedback line of Figure 1.
MEASUREMENT AND CONTROL OF WELD POOL PENETRATION

Weld Vision System

Electro-optical techniques are used for the indirect measurement of penetration. These techniques are attractive because contact with the workpiece can be avoided and problems with physical wear and heat transfer to the sensor are thereby minimized. Such systems can also accommodate a variety of weld joints and bead geometries by use of adaptive image processing, etc. On the negative side, one can expect optical systems to be relatively intolerant of smoke or aerosol generation at the weld site and to liquid metal spatter. However, a variety of schemes are available to deal with these problems, including the use of a purge gas in the optical path, the use of mechanical shields, and positioning optical surfaces at an appropriate distance to achieve significant cooling of incident spatter.

In 1982, the INEL began exploratory work on the use of machine vision for electric arc welding. Earlier work revealed that the welding arc light is a very severe impediment to formation of good imagery and must be suppressed and/or replaced by illumination from an external light source. Figure 3 is a simple schematic of the experimental arrangement. The goal is to obtain enough peak optical power from the xenon flash lamp to overwhelm the welding arc emission during the brief 2-3 μs interval of the flash. A video camera system equipped with an image intensifier tube is used in a time-gated mode as a very high speed electro-optical shutter. The shutter is synchronized with the flash lamp and therefore acts to accept most of the flash energy reflected from the weld site, but at the same time accepts only a very small fraction of the continuous emission from the welding arc. The synchronized flash and shutter are driven at 30 pulses per second to yield a single flash exposure per video frame.

These early experiments were done with a small gas tungsten arc (GTA) welder operating at 60 to 70 A arc current on stainless steel. Using this system, the visibility through the arc is greatly enhanced, with the weld pool in black contrasting sharply with the solid material in grey. The pool appears black because the liquid metal surface is a very good specular reflector and the xenon flash energy, which is incident from the right, is reflected directionally to the left and not into the video camera field of view. The important result is a quality, high-contrast, video image that can be interfaced with a digital image processing system to automatically characterize the welding pool geometry and geometrical relationship between pool, electrode, and seam.

To improve system performance, a variety of pulsed laser systems were studied as candidates to replace xenon flash illumination, and the nitrogen laser was selected

Welding Video Technique

![Welding Video Technique Diagram](image)

Figure 3. Schematic of weld vision system.
for its relatively high reliability and reasonable price. A pulsed laser is attractive for this application because of the very intense peak optical power levels achievable, the single-wavelength emission which allows the use of narrow band spectral filtering, and the very good focusing characteristics of the beam, which allows transfer of the light energy to the torch via optical fiber.

The nitrogen laser radiates in the near ultraviolet region at 337 nm which does not overlap significant emission lines of the shield gases or alloy materials used in arc welding. This wavelength seems to be suitable in this respect for all welding experiments done to date. Figure 4 shows GMA welding of heavy-section aluminum alloy plate. In this process the shield gas is pure argon and aluminum wire is used as a consumable electrode in place of the tungsten electrode used in the GTA process. The welding current was ~400 A and the weld bead was ~29 mm wide in a flat-bottom V groove. The typical video picture without laser or shuttering is shown in Figure 4a. Some evidence of the wire electrode can be seen below the rim of the gas cup, but little detail can be seen in the groove. The bright elliptical area represents the arc light surrounded by a large depression in the welding pool. A great deal more detail can be seen with laser illumination (Figure 4b).

The camera is presently being used to measure weld pool widths and help understand factors governing weld pool surface characteristics. A technique is being developed that will allow the topographical features of a molten weld pool to be measured. This method capitalizes on the fact that the weld pool is essentially a mirror surface which reflects a distorted image of the weld pool surface. The amount of distortion (topographical change) is measured by comparing the distorted reflection of a known pattern of holes in an aperture array with the undistorted image reflected from a flat mirror. Images of the aperture plate pattern are recorded on video tape, using the weld pool surface as a mirror. This pool has a complex and continuously changing contour, which is due to phenomena such as convection currents and surface tension gradients. The data are then analyzed using a complex mathematical formulation assessing the weld pool topography. Presently, this work is limited to gas tungsten arc welding (GTAW).

Ultrasonic Sensing of Weld Pool Penetration

In contrast to sensing the surface of the weld pool using electro-optical methods, ultrasonic sensing uses sound waves to "look" at the interior of opaque materials. A transducer operating in the pulse-echo mode converts electrical signals into high-frequency sound waves that travel to the area of the weld pool. When the sound wave strikes the metal/molten metal interface, a portion of the energy in the wave is reflected. Some of the reflected sound energy may find its way back to the transducer, which converts the reflected sound into an electrical signal that can be observed on an oscilloscope or digitized by an analog-to-digital converter for computer analysis. This ultrasonic echo provides information about the location of the interface.

The current welding research involves positioning a transducer on the topside of a weld sample, ultrasonically sensing defects 330 mm behind the welding torch in the solidified weld metal, and analyzing the digitized data using pattern recognition techniques. The real-time monitoring system for detecting the molten weld pool uses a 5 MHz contact transducer mounted on various lucite wedges to generate refracted ultrasonic beams in carbon steel. A fixture design allows the transducer, mounted on the lucite wedge, to move along the plate parallel to the weld preparation in alignment with the welding torch (Figure 5). Carbon steel 25.4 mm thick is used in two weld geometries—a single bevel V-groove having a 30° included angle, a 4.76 mm root opening, and a 6.35 mm backup bar; and a V-groove having a 60° included angle, a 4.76 mm root opening, and a 6.35 mm backup bar. Ultrasonic data is
collected using either a video system which records the real-time ultrasonic signal displayed on the oscilloscope or a DEC LSI 11-23 computer to digitize and store the amplified ultrasonic signal.

Using the transducer fixture and the encoder mounted on the side beam welder, alignment is established between the torch and transducer. Data have been acquired using various lucite wedges which provide different angles of refraction and modes of sound propagation in the weld sample to sense the molten weld pool. A 45° refracted longitudinal sound wave allows sensing of the molten weld pool interior since longitudinal sound waves propagate in the liquid metal. This sensing technique can detect a good weld, porosity in the pool, and lack of side wall penetration on the side of the weld nearest the transducer. A 60° refracted shear wave on a 30° bevel plate allows sensing of the molten-solid interface (shear waves are not
supported in liquid) and amount of side wall penetration. The 60° refracted angle is not optimum for sound reflection at the molten/solid interface; however, it is valuable in determining if there is full side wall penetration.

An evaluation process is currently under way using the 45° refracted shear wave data. Reflectors seen before, during, and after welding are evaluated using machined samples to simulate these three unique points of the welding process. Initial correlations between the machined samples and the acquired data indicate that the geometries of the forming molten/solid interfaces can be detected and grouped into several basic geometry types. This information provides the potential of monitoring the interface and providing control input which can assure an optimum interface geometry throughout a welding pass.

CONCLUSION

Process sensing and control of automated welding is possible for the selected parameters discussed above. The success of this integrated control scheme can result in improved weld quality, increased productivity, and energy cost savings in appropriate welding applications.

ACKNOWLEDGMENTS

This work is supported by the U. S. Department of Energy, Office of Energy Research, Office of Basic Engineering Sciences under DOE Contract No. DE-AC07-76ID01570. Appreciation is expressed to U. S. Wallace, C. L. Shull, and C. Stander for technical assistance.

REFERENCES

VISIBLE LIGHT EMISSIONS DURING GAS TUNGSTEN ARC WELDING AND ITS APPLICATION TO WELD IMAGE IMPROVEMENT

E. W. KIM, C. ALLEMAND, T. W. EAGAR,
Dept. of Materials Science and Engineering
Massachusetts Institute of Technology

ABSTRACT

An experimental study was carried out to map the light emissions from gas tungsten arc using 24 combinations of weld parameters. Data were collected with a computer interfaced double monochromatic imaging system. The intensity scale was calibrated throughout the wavelength range from 488 nm to 730 nm using a spectral radiance standard lamp which is traceable to NBS standards. Lines were identified by element and wavelength with 0.065 nm wavelength resolution.

The effect of intensity calibration was found to be very important in terms of comparing spectral maps and intensities. The emissions were found to be dramatically different with different shielding gases, welding current and base metals. Spectral windows were also determined. An improved image of the weld puddle can be obtained by operating within these spectral windows. Two dimensional distributions of major elements in an arc plasma are also presented and the anode spot on the molten weld surface was observed.

INTRODUCTION

Improvement of automated welding will require development of new sensor systems. As the welding arc is a harsh environment, non contacting sensor devices are preferred. A number of investigators have developed methods for viewing the weld pool in order to control seam tracking and weld bead width; however, each of these systems must exclude the arc light in order to obtain a clear image of the weld pool [1-4]. It would be desirable to map the spectra on a variety of base plates in order to select the windows in the spectrum which are relatively free of arc light. Arc light emissions also have potential as a welding information source. The emission changes with weld parameters and responds well into the kilohertz range [5]. In most cases, the intensities from different investigations and even from the same investigation are not comparable because of a failure to calibrate the measurement systems [5-8].

The major purpose of the present work is to spectrally map light emissions in the range of 488 nm to 730 nm during gas tungsten arc welding, and to correlate
changes in the spectra to changes in the welding parameters. The intensities are calibrated so that all intensities are measured on the same scale. Mapping the emission with various combinations of weld parameters provides comparison of the emission characteristics. Further, this mapping defines spectrally blind regions (of minimum spectral intensity). These regions can be a basis from which to choose windows in the spectra where external sensors will have the least disturbance from welding arc emissions. The spectral distribution of the major spectral lines were also photographed based on the results of this spectral mapping.

**EXPERIMENTAL WORK**

The spectra obtained in this study were made with the equipment shown in fig.1 [9]. The image comprised the molten surface, the arc column and the cathode tip. The imaging double-monochromator used in this study has a 500 mm focal length and a reflection grating of 1180 grooves/mm.

![Fig.1: Schematic Arrangement of Spectral Data Analysis System](image)

A schematic drawing of this instrument in fig.2 shows the light path through the instrument. The spectral resolution used is 0.065 nm with 20% maximum error in determining the peak intensity. The rejection ratio of the monochromator as a band pass filter was also measured to be better than $10^6$.

The intensities were calibrated by comparison with a lamp whose radiance is traceable to NBS standards. This lamp was used as the light source at the position of the arc source and output intensities were taken through the entire opto-electronic system. The ratio of this output to the lamp radiance was used as an intensity calibration factor. This factor was used to normalize the intensity over the entire wavelength range studied; however, since both the entrance slit size and the source position may be changed during the measurement to obtain the best signal condition from the welding arc, the calibrated spectral intensities are only relative. Nevertheless, the relative intensities at any two wavelengths are linearly
proportional to the peak heights.

![Diagram of light path in a double-konochromator](image)

**Fig. 2:** Light Path in the Double-Konochromator

AISI 304 stainless steel, Ti-6Al-4V, 5083 Aluminum alloy and copper plate were used as base metals. Three different welding currents, 100, 150 and 200 amperes were used with electrode negative polarity (DCEN) and an 8 mm arc length. Welding grade argon and helium were used as shielding gases at flow rates of 20 cfh for Ar and 25 cfh for He. Spectra were taken at each of these 24 experimental combinations.

**RESULTS AND DISCUSSION**

The accuracy in determining wavelengths are listed in table I. The examples of spectral maps for the case of AISI 304 stainless steel are shown from figure 3) to 6). The spectra at wavelengths longer than 690 nm usually consist only of strong Ar lines, namely Ar I 6965.43, 7067.23, 7147.04, 7270.94. These were omitted from the spectral map. Some lines which saturated the upper limit of the data collecting system are marked with upward arrows. The relative intensities in the wavelength tables are somewhat different from those of the experimental results. By and large, strong tabulated lines match to the test results but differences often occur when the intensity level of a spectral line is low. This may be due to the interaction effects of each element in the arc.

Since metal vapors may dominate the temperature profile and the electrical properties of the arc, it can be seen that a change in the composition of the work piece can dramatically change the nature of the welding arc [8,12]. The major lines from the He arc on a water cooled copper plate are Cu I and He I. In the case of the Ar arc, Ar I and Ar II lines are dominant. This is due to the fact that He gas
### TABLE I: Band Pass of Imaging Double Monochromator.

<table>
<thead>
<tr>
<th>slit width (micron)</th>
<th>calculated FW_HI (nm)</th>
<th>measured FW_HI (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.763</td>
<td>0.763</td>
</tr>
<tr>
<td>250</td>
<td>0.381</td>
<td>0.346</td>
</tr>
<tr>
<td>100</td>
<td>0.153</td>
<td>0.140</td>
</tr>
<tr>
<td>50</td>
<td>0.076</td>
<td>0.074</td>
</tr>
<tr>
<td>40</td>
<td>0.061</td>
<td>0.056</td>
</tr>
<tr>
<td>30</td>
<td>0.046</td>
<td>0.055</td>
</tr>
<tr>
<td>20</td>
<td>0.031</td>
<td>0.055</td>
</tr>
</tbody>
</table>

FW_HI = full width at half intensity

---

**Fig. 3:** Effect of Current on 4880 Å - 7300 Å Spectra of He Arc on AISI 304 S.S.

**Fig. 4:** Effect of Current on 4880 Å - 6950 Å Spectra of Ar Arc on AISI 304 S.S.

**Fig. 5:** Effect of Current on Spectra of He Arc on AISI 304 Stainless Steel, 5130 Å - 5268 Å

**Fig. 6:** Effect of Current on Spectra of Ar Arc on AISI 304 Stainless Steel, 5130 Å - 5268 Å
liberates more heat than Ar gas. This can be explained by the fact that the arc voltage of He shielding is much higher than that of Ar shielding due to the higher ionization potential of He. This trend was consistent with each experiment performed. The major metallic lines observed in AISI 304 stainless were Cr I and Fe I. The dominant metallic lines from the 5083 aluminum alloy experiment are shown to be Mg I and Al II. Mg lines saturated the upper limit of the data collecting system. The major lines from this experiment are Ti, Ti I and Al II. Thermodynamic calculations made by Block-Bolten and Eager and Cobine et al. are consistent with these observations [13-15].

The data can also be presented in finer detail as shown in fig. 5) and 6). An atlas of light emissions during arc welding with various combinations of welding parameters has been produced [12]. It is seen that the uncalibrated intensity ratio between two lines compared with the calibrated ratio can be in error by as much as 15 times. This strong effect of calibration on relative intensities explains some of the difficulty in comparing other spectra that have been presented in the literature.

It is clear that most of the lines appear at all current levels on the same material but with quite different relative intensities. Some of the relative intensity changes are quantified in figure 7 and 8.

![Fig. 7: Effect of Shielding Gas and Current on Emission Intensity of Cr I and Cr I + Fe I Lines from AISI 304 S.S.](image)

![Fig. 8: Effect of Shielding Gas and Current on Emission Intensity of Fe II, Fe I, W Lines from AISI 304 S.S.](image)

![Fig. 9: Image of Ar Arc on 304 Stainless Steel](image)

![Fig. 10: Anode Spot Mode on 304 Stainless Steel at Cr+Fe 520.6+1.4 nm](image)
The intensities from neutral lines increase with Ar shielding but decrease after reaching a maximum with He shielding. On the contrary, the intensity of singly ionized lines in figure 8 shows a consistent decrease. This may be due to a transition from a singly ionized Fe ion changing to a secondary ion emitting light at a wavelength which is not within the range concerned in this research. Essentially, the plasma may become more optically thick at this wavelength as the current is increased [16]. It is also seen that the intensity of the tungsten line from the electrode increases with current when using He shielding.

The background intensity with Ar shielding at 200 amperes is approximately 20 times higher than that with He shielding. One possible explanation for this is the higher electrical conductivity and the lower thermal conductivity of the Ar plasma compared to the He plasma [8]. Another possible explanation may be the free to bound electron transition of the singly ionized Ar atoms due to their lower ionization potential compared to that of He. Although the background is lower, the peaks are higher with He for any line which appears in both shielding gases. This may be due to the higher arc power of the He gas shielding caused by the higher total arc potential drop, or it may be due to the fact that He is optically thinner than Ar.

The base metal also affects the intensities of some lines. For example, the intensities of the same line, 5875.62 He I, differs greatly according to the composition of the base metal. This probably results from the interactions of each vapor or ion species in the arc plasma, again confirming the fact that the metal vapor can dominate the properties of the arc. The property change of the plasma due to the presence of metal vapors requires a more rigorous study in the future.

As stated in the introduction, the spectrally blind regions can be found easily. By superimposing eight spectra from four different base materials in one graph, windows were found to be at 579 nm – 587 nm, 589 nm – 601 nm, and 644 nm – 663 nm for these four base metals. Windows for fine wavelength scales can also be found from graphs similar to fig.5 and 6 [12].

Photographs were taken with 652 nm interference band pass filter which has a 10 nm band pass FWHM. These are compared with the photographs taken with neutral density (ND) filters. As shown in figure 9, better images of a molten surfaces were obtained. As stated previously, Ar shielding has a higher continuum emission than He, hence, the arc light is still present in fig.9 (a). This emission is concentrated in the vicinity of the tungsten electrode. Figure 10 is image of Cr+Fe which was made with the double monochromator using its 1.4 nm band pass FWHM. These are heavily concentrated just above the molten pool. This observation is in accordance with the results of population density measurements in an arc plasma [17]. Figure 10 (a) shows a spot at the center. This bright spot is believed to be an anode spot where the current density and temperature is higher than other locations on the weld pool surface.

It was found that the shape of the anode spot changes with increasing current from a spot to a ring. Figure 10 (b) shows this. The center ring is believed to be the locus of the anode spot and the inverted cone is the reflected image of the arc column. The fusion boundary is marked with a white curve. The transition occurs at around 150 amperes. The ring shape is believed to be related to the depressed molten anode surface [18]. At high currents above 150 amperes, the anode spot is thought to move around the ridge of the depressed pool surface with a non-random motion. The video observation showed a rotating motion of the ring. This will affect the distribution of the metallic elements in an arc plasma and thus the heat input mode to the work piece. A ring shaped heat input model should be considered and the electromagnetically driven convection flow in the weld pool will be changed in comparison with typical mathematical models of arc welding. The rotating spot along the ridge may drag the molten metal and work as an asymmetric driving force for the weld pool cir-
Spectral maps for 24 combinations of weld parameters were made in the wavelength range from 488 nm to 730 nm using a calibrated intensity scale with 0.065 nm wavelength resolution. Lines can be identified by species and wavelengths within this resolution. The uncalibrated ratio between two lines could be in error by as much as 15 times.

The major metallic lines identified were Cr I and Fe I for AISI 304 stainless steel, Cu I for copper, Ti, Ti I, Al II for Ti-6Al-4V and Mg I and Al II for 5083 aluminum alloy. The lines from the shielding gases were found to be very strong. Ar gas shielding at 200 ampere arc current showed approximately 20 times higher background emission level compared to the He gas shielding, but the peak intensities were generally higher with He gas shielding than with the Ar gas shielding. The intensity changes at different current levels were found to vary from element to element and were also found to be affected by the shielding gas. It was found that windows are present at 579 nm - 587 nm, 589 nm - 601 nm, 644 nm - 663 nm.

The anode spot was observed to change its shape from a spot to a ring with increasing current. It is believed that this may be due to movement of the anode spot around the outer ridge of the depressed weld pool surface. Hence, it is seen that anode spot motion on molten weld pools at high welding current (above 150 amperes) is non-random. This could have an important effect both on the heat distribution of non-planer molten weld pools and on the electromagnetically driven convection within the weld pool.

ACKNOWLEDGEMENTS

The authors are appreciative of support for this work from the Department of Energy, under Contract DE-FG02-85ER13331.

REFERENCES


183
ABSTRACT

The task of controlling welding is one that involves a multivariable and multi-energy domain process. Most work in this area has been directed at the development of sensing methods and related control systems for single variable, low bandwidth operation. In this paper the system dynamics of the process is presented in a pair of comprehensive models for the control of pool geometry and for the control of thermal history during Gas Metal Arc welding. Through physically based modeling of these two aspects of the process, the full non-linear models are developed. In the case of geometry, the model is derived from known behavior of the arc, approximations as to pool development and a simple heat transfer analysis of the weldment. The thermal problem is developed through a numerical analysis that includes accurate representations of material properties and heat transfer mechanisms involved. In this case, experimental verification was obtained. In both cases the non-linear models are reduced to equivalent linear transfer functions, and these are shown to be non-stationary, with a strong dependence on the operating point. These findings have strong implications for the method of control of multivariable systems, and their ultimate performance.

INTRODUCTION

The ultimate goal of welding is to join two or more materials with sufficient strength and fracture toughness for the assigned task without excessive distortion of the base materials. These properties are the integration of many individual aspects of the welding process, yet, the task of welding control is to achieve high quality by controlling the melting and resolidification processes involved. With this indirect control authority, it is imperative that a comprehensive control strategy be pursued to prevent a fragmented, redundant, and ultimately low performance solution from resulting. With respect to quantities that can be controlled during the welding processes, the dominant ones are: fusion zone geometry, heat affected zone properties, and gross weldment thermally induced stresses. (Transient imperfections such as porosity or inclusion are of great importance but are assumed at this point to be best addressed by material and weld schedule selection).

This paper explores the system dynamics for the weld geometry and thermally activated material properties for Gas Metal Arc Welding (GMAW). The models demonstrate the truly coupled and non-linear nature of the welding process. Equivalent linear models are derived in an attempt to provide models for controller design. These transfer functions indicate the non-stationary nature of the process, and point to the
need for non-linear controller design.

CONTROL OBJECTIVES

The goal of most control systems is to achieve stable regulation (i.e. setpoint maintenance) when the process under control changes or is subjected to unpredictable outside influences. This description certainly captures the essence of welding, but a more concise objective can be stated. In this research, we are concentrating on achieving simultaneous regulation of five quantities, three related to fusion zone geometry and two related to heat affected zone properties.

The geometry can be characterized by the weld surface width, maximum depth and the area of reinforcement provided by the filler material. These quantities determine the basic mechanical properties of the weld and are key to insuring the basic stress handling capabilities of the joint. The solid material properties are determined (in differing ways depending upon the material in use) by the extent of some critical isotherm (heat affected zone width) and the rate of cooling at some critical temperature. These control objectives, constitute a five output control system (see Fig. 1), and since each of these quantities is affected by each of the available controls on a GMAW torch, a fully coupled control strategy must be adopted.

To pursue such a study, a model of the input-output characteristics of this "block" must be understood, both to identify what inputs to the process (e.g. heat rate, wire feed, etc) will be needed, and to permit a range of controller designs to be explored so as to optimize performance and assess the impact of various feasible feedback measurement schemes. The remainder of this paper will detail models necessary for the characterization of welding to permit full multivariable control of the process. While ultimately there must be a single, five input, five output model as described above, two separate analyses are presented, one for geometry and one for thermal effect. However, the coupling of the two will be obvious and is reviewed at the end.

FUSION ZONE GEOMETRY CONTROL

Geometry control of the GMA welding process requires control of three output variables, weld bead width, w, bead depth, d, and reinforcement height, h. There are three control variables available for regulation of the output variables; open circuit voltage, $E_o$, wire feedrate, $f$, and torch speed, $v$. The dynamic relationship between the control variables and the output variables is expressed by a control model.

The basic system to be modelled is shown in Fig. 2. Of concern to this analysis is the power supply - arc - electrode interaction, the dynamics of the weld melting process, relationships between the weld pool width and depth, and the dynamics of the servos used to control the electrode wire feedrate and torch travel speed.

A non-linear model of the welding system has been developed in [1]. The form of the model is:

\[
\begin{align*}
\frac{dE}{dt} &= \left(\frac{2(1+\lambda_m)}{\eta_2(E_o+\eta_1)}\right) \left[\begin{array}{c}
\eta_1(E_o+\eta_1) \\
\eta_2(E_o+\eta_1)
\end{array}\right] \\
\frac{df}{dt} &= \frac{-f + f_o}{f_f} \\
\frac{dv}{dt} &= \frac{-v + v_o}{v_v}
\end{align*}
\]

185
where \( l, f, w, \) and \( v \) are the four states arc length, wire feedrate, bead width and torch velocity respectively and \( E, f, \) and \( v \) are the control inputs open circuit voltage, wire feedrate and torch velocity. Other variables are as follows:

- \( R \) resistivity of the wire
- \( h \) contact tip to workpiece distance
- \( I \) current
- \( \eta_1 \) portion of the heat generated by the arc voltage drop which is transferred to the wire
- \( F_t \) voltage drop across the anode and cathode
- \( g \) slope of the arc voltage across the arc
- \( Q_{r1} \) heat required to melt the wire
- \( I_m \) maximum arc length
- \( Q_{r2} \) heat required to melt the base material
- \( \eta_2 \) sloping factor of the power supply
- \( K \) efficiency factor
- \( K_v \) conductivity of the metal
- \( A_t \) area of the solid-liquid interface, \( a \)
- \( \eta \) function relating temperature gradient to torch speed
- \( T \) time constant for the torch velocity servo

The three outputs are width, \( w \), depth, \( d \), and reinforcement height, \( R_h \) where

\[
d = \frac{w}{1 + 1_m} \quad R_h = \frac{4A_v}{\pi w w}
\]

and \( A_v \) is the cross-sectional area of the electrode. These equations constitute a physically based model which is modified from similar forms in [2]-[4]. The model has not been verified experimentally and some of the variables are not well known. Thus the model does not yield exact results, but does provide vital information about the form of the system response and the dynamic interaction between inputs and outputs.

This simple model, then, has four state equations that describe the relationship between the physical welding process including actuator dynamics and the control inputs. There are also three equations which show the relationship between the four states and the three outputs. With this model we have an indication of how the outputs will respond to changes in the inputs, and it shows very clearly the non-linear nature of the process and also the strong coupling between the states.

Figures 3-8 show the response of the non-linear system to step changes in the control inputs. The plots show the change from a nominal steady-state condition which is \( f=8.33 \) in/s, \( l=0.312 \) in, \( w=0.5628 \) in, \( v=0.15 \) in/s, \( d=0.37 \) in, \( R_h=0.121 \) in, \( E=32 \) v. The non-linear effects can be seen in the different responses to positive and negative steps, especially for \( R_h \) in Figures 7 and 8. The coupling is obvious since each output responds to each input. The model predicts a very rapid initial change in the outputs and then a slower drift to a steady-state condition. This is a result of the lumped-parameter nature of the model. The weld pool under the arc is a continuous, rather than lumped-parameter, system which tends to filter the responses, so the rapid initial change probably can not be observed in an actual bead. The changes will tend to be averaged over the whole weld pool.

Notice also that the control power available from \( E \) is much less than that available from \( F \) and \( V \). This is because a unit change in \( E \) represents about one-fourth of the full control range. This limits the control space of the welding system and the amount of decoupling which is possible. There are, however, other control options. Standoff, or contact tip to workpiece distance, is another variable which
could be manipulated as an input. Current pulsing can also be used to increase control power or enlarge the control space.

The non-linear geometry model can be linearized about an operating point and expressed in transfer function notation. The linearized response of the geometry variables \( Y \) to the inputs \( U \) may be written in the form of a transfer function matrix \( G \) as follows:

\[
Y(s) = G(s)U(s)
\]

with

\[
Y(s) = \begin{bmatrix} W(s) \\ D(s) \\ R_h(s) \end{bmatrix}, \quad U(s) = \begin{bmatrix} F_c(s) \\ E_0(s) \\ V_c(s) \end{bmatrix}, \quad G(s) = \begin{bmatrix} -2.06F_2^2T_5(T_2+1) & -K_1F_2T_3(T_2+1) & -10.28F_3 \\ F_c & E_0 & V_c \\ \tau_7T_6(T_2+1)(T_3+1) & \tau_6(T_2+1)(T_3+1) & \tau_{10}(s+20)(T_3+1) \\ 0.145T_2^2T_3(T_2+1)(T_3+1) & K_1F_2T_3(T_2+1) & K_3T_3^2(T_2+1)(T_3+1) \\ F_c & E_0 & V_c \\ \tau_7T_6(T_2+1)(T_3+1) & \tau_6(T_2+1)(T_3+1) & \tau_{10}(s+20)(T_3+1) \end{bmatrix}
\]

where \( W(s) \) and \( D(s) \) and \( R_h(s) \) denote deviations from the nominal conditions and the elementary transfer functions of \( G(s) \) are found by a direct Taylor Series linearization of the state equations. These are given by:

\[
\begin{align*}
W(s) &= \frac{-2.06F_2^2T_5(T_2+1)}{F_c} \\
D(s) &= \frac{-K_1F_2T_3(T_2+1)}{E_0} \\
R_h(s) &= \frac{0.145T_2^2T_3(T_2+1)(T_3+1)}{F_c}
\end{align*}
\]

\[
\begin{align*}
W(s) &= \frac{K_1F_2T_3(T_2+1)}{E_0} \\
D(s) &= \frac{K_3T_3^2(T_2+1)(T_3+1)}{V_c} \\
R_h(s) &= \frac{K_3T_3^2(T_10s+1)}{V_c}
\end{align*}
\]

The \( K \)'s and \( r \)'s are listed in Table 1 for the nominal condition stated above and for the step changes in control variables corresponding to Figures 3-8. Note the high variability of any given parameter (gains \( K \) and time constants \( r \)) at various different operating conditions.

As a test of the sensitivity of the operating point used in the linearization, the linear model response to the input of Fig. 3 (+\( F_c \)) was examined. When a linearization point equal to the steady state values was assumed, the correspondence between linear and non-linear models was acceptable. However, if the nominal conditions were used for linearization, the resulting linear model response deviated considerably from the non-linear. This effect is shown with the dashed line in Fig. 3.

**THERMAL PROPERTY CONTROL**

The following analysis focuses on the thermal control problem, i.e. thermal variables will be used to control the material microstructure characteristics of the weld joint. Before selecting the suitable control inputs, the appropriate quantities characterizing the state of the process and determining the important metallurgical fea...
tures of the final structure must be specified. As covered in detail in [5], these will be the weld nugget cross section area, NS, the heat affected zone width, HZ, and the centerline cooling rate at a critical temperature, CR. Roughly speaking, the weld nugget cross section area, defined by the solidus isotherm $T_B$, represents a collective measure of the extent of solidification defects in the weld bead, and may also be important in controlling the dilution of the base material with filler material. The heat affected zone width, defined by an envelope isotherm $T_H$, may indicate the extent of weak areas adjacent to the bead (such as the recrystallization zone), or areas in which some undesirable phase nucleates (such as the sensitization zone in which Cr$_7$C precipitates), or areas vulnerable to contamination because of increased reactivity with the environment. Finally the centerline cooling rate at the critical temperature $T_c$ gives a measure of the effects of thermal stresses (such as hot cracking or the extent of nucleation of undesirable phases such as martensite). The degree of importance of each of the three features of the temperature fields is different for different types of welding material and is reviewed in [5] for a wide range of materials.

The steady-state model

Steady state relationships for the three variables NS, HZ, and CR were derived by Doumanidis and Hardt [6]. (As discussed in [6], the independent control of CR requires more than a single heat source, and in this case a second torch of heat input $Q_2$ is assumed, and the ration $S = Q_1/Q_2$ becomes the third input to the system.) These can be combined to form a complete steady-state model, where the dependence of the variables on the three inputs $Q_1$, $v$ and $S$ is explicitly expressed:

$$NS = C_1 \left[ \frac{Q_1}{v} \right]^{n_a}$$
$$HZ = C_2 \left[ \frac{Q_1}{v} \right]^{1/n_b} \cdot \left[ \frac{1}{v} \right]^{1/n_a} \cdot \left[ \frac{1}{Q_1} \right]^{1/n_c}$$
$$CR = C_3 \left[ \frac{v}{Q_1} \right]^{n_b} \cdot \left( 1 - \beta S \right)$$

This is a steady-state, partially decoupled model of the dependence of NS, HZ and CR on $Q_1$, $v$ and $S$. Note that $Q_1$ is any measure of the heat input provided by the primary source, such as voltage (if the current is kept constant) or current (if the voltage is kept constant) or the product of voltage and current, since the conversion factors are incorporated in the constants $c$. As it is shown in [6] the constants $C_1$, $C_2$ and $C_3$ and the exponents $n_a$ and $n_b$ can be determined by information available in-process, while the exponent $n_c$ and the coefficient $\beta$ must be known by previous experiments.

Transient Modelling - Numerical Simulation

In order to model the dynamic dependence of NS, HZ and CR on the selected inputs $Q_1$, $v$ and $S$ a solution of the time-dependent temperature field described by the transient conduction model, which is based on the integration of the appropriate Green's function [7], turns out to be impractical and limited in scope. Instead, a versatile and general-purpose numerical simulation of the temperature field of the multitorch welding process was developed. The computer program, which integrates the unsteady heat transfer equations in discrete time and space elements by using an explicit backward finite difference formulation.

The geometrical configuration of the simulation is that of butt welding two identical, infinitely long and wide plates with finite thickness (2-1/2-D configuration). This case covers with acceptable accuracy the finite geometries where the
width of each plate and the distance from a longitudinal edge at the beginning of welding are greater than ten times the plate thickness [8]. For even more restricted geometries where this condition is not fulfilled, the simulation is still applicable with minor modifications. Another requirement is the symmetry of the heat flow problem about the center plane, which permits studying the temperature field in one of the plates only. For asymmetric geometry or other heat flow conditions, there is a provision for heat exchange at the interface of the plates, which will permit concurrent simulation of the temperature field in different adjoining plates.

The geometry near the torches is simulated by two rectangular constant-spacing grids moving together with the torches. A coarse grid serves in the study of the temperature field at remote areas (HZ, CR) while a finer grid ensures the required accuracy near the weld pool (NS). The configuration of the process is a multitorch one, with all of the torches aligned on the centerline. The number and distances between the torches are user-defined. The heat input from each torch is spatially distributed according to a Gaussian distribution, which is alterable to any arbitrary form, permitting even, continuous distributions along the centerline. The heat outflow from the welded pieces includes convection and radiation, with different coefficients for the two sides of the pieces.

The properties of the material are temperature-dependent, with the corresponding function defined by the user for each material. These include solidification-fusion latent heat to take into account the thermal effects of this transformation between the solidus and liquidus temperature in the general case of an alloy. The weld pool heat transfer consists of equivalent conduction mechanisms only. The equivalent conduction coefficient used for the weld pool material can be adjusted to make any characteristic feature of the pool (such as NS) match experimentally determined values.

Experimental evaluation of the model was performed using single torch GMA butt welding of thin mild steel plates, with 1/8 in thickness. The wire feed and gas (Ar) flow were constant, and no preheat was used. The heat affected zone was enveloped by an assumed recrystallization isotherm $T_h = 826 \text{ K}$, and the cooling rate was measured at the austenitization temperature $T_c = 996 \text{ K}$.

The on-line measurements consisted of scanning the time-dependent temperature field on the top and bottom surface of the plates, using an infrared radiation camera. The bead geometry was studied by successive transverse sectioning and microscopic observation of the bead cross section. The experimental calibration of the parameters was performed at a set of inputs producing outputs approximately in the middle of the specified or desirable range (usually limited by unacceptable fusion penetration or lack of fusion and excessive porosity). These "nominal" conditions were:

\[
\begin{align*}
Q_1^* &= 2500 \text{ W} \\
\nu^* &= 5 \text{ mm/s} \\
Q_2^* &= 0.0 \text{ W}
\end{align*}
\]

\[\text{corresponding to} \quad NS^* = 3.30 \text{ mm}^2 \quad HZ^* = 5.39 \text{ mm} \quad CR^* = -82.2 \text{ K/s} \]

Fig. 9 presents the experimentally observed values of the outputs during the transient from zero to the nominal conditions, with the solid line representing the response of the simulation to the same set of inputs. The model parameters were calibrated so that this response matches the experimental data in an approximate ITAE sense (yielding an arc efficiency of 0.869, a heat distribution radius of 1.71 mm and an equivalent weld pool conduction factor of 3.57).

Having calibrated the dynamic model, the dynamics of the welding process can now be studied both by experiment and by simulation. This is done by starting at the nominal conditions and obtaining the transient response of the welding system to steps in the inputs $Q_1$, $\nu$, and $Q_2$. The steps applied to each input go approximately half the way between the nominal conditions and the most extreme values of inputs that the sys-
tem is expected to handle in practice. The simulations performed are listed below:

<table>
<thead>
<tr>
<th>Step Input</th>
<th>Initial Value</th>
<th>Final Value</th>
<th>Fig #</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Positive step to $Q_1$</td>
<td>$Q_1 = -2500W$</td>
<td>$1.2*Q_1 = 3000W$</td>
<td>Fig 10</td>
</tr>
<tr>
<td>b. Negative step to $Q_1$</td>
<td>$Q_1 = -2500W$</td>
<td>$0.8*Q_1 = 2000W$</td>
<td>not shown</td>
</tr>
<tr>
<td>c. Positive step to $v$</td>
<td>$v = -5mm/s$</td>
<td>$1.2*v = -6mm/s$</td>
<td>Fig 11</td>
</tr>
<tr>
<td>d. Negative step to $v$</td>
<td>$v = -5mm/s$</td>
<td>$0.8*v = -4mm/s$</td>
<td>not shown</td>
</tr>
<tr>
<td>e. Positive step to $Q_2$</td>
<td>$Q_2 = 0.0 W$</td>
<td>$Q_2 = -250 W$</td>
<td>Fig 12</td>
</tr>
<tr>
<td>f. Negative step to $Q_2$</td>
<td>$Q_2 = 0.0 W$</td>
<td>$Q_2 = -250 W$</td>
<td>not shown</td>
</tr>
</tbody>
</table>

Thus, for example, Fig 10 presents the measured and calculated output response after a step increase of the heat input $Q_1$ by 20%. The response of the calibrated simulation is indicated by the dashed line, which sometimes is at a small steady-state error with respect to the experimental data. These can be fitted better in an approximate ITAE sense by just a slight adjustment of the calibration parameters, in which case the response of the simulation is shown with the solid line. Finally the dot-dashed line corresponds to the response of a lumped parameter, linearized model, which is discussed later. The same notation is used in Figs 10-12.

Fig 12 illustrates the responses of the outputs to a positive step applied to the heat input $Q_2$ of a secondary torch positioned 10mm behind the main torch. Since an experiment using the second torch was not performed, the predicted responses may include small quantitative inaccuracies as before.

The non-linear nature of the system is illustrated by the different response to inputs of different magnitude, i.e. the dependence of the above step responses of all the outputs on the value of the step applied to each of the inputs. Indeed, the steady-state deviations of the outputs from the nominal conditions after the transient are markedly different for positive and negative steps of the same size applied to the inputs and there is also an apparent dependence in the settling time. Also, the dynamic dependences are of higher order, as it is clear from the often irregular and non-smooth nature of the step responses.

LINEARIZED THERMAL SYSTEMS DYNAMICS

The dynamic model can be linearized in the neighborhood of the nominal conditions for limited ranges of variation of the inputs, and described in terms of low-order transfer functions. The form and parameters of the latter must be selected so that the responses match as closely as possible the previous experimental data (or the non-linear simulation response in the case of the secondary heat input $Q_2$). In the following analysis the values of the welding inputs and outputs again refer to deviations from the nominal conditions.

The response of the bead cross section area $NS$ to steps in either $Q_1$ or $v$ can be approximated by the expected over-damped second order behavior, in which one pole clearly dominates over the other, so that $NS$ may be adequately modelled by a first order transfer function with respect to either $Q_1$ or $v$ [4]. As it appears in Fig. 12, $NS$ does not respond to small enough steps of $Q_2$, i.e. it is decoupled from the effect of the secondary torch. Consequently, the proposed transfer functions are of the form:

\[
\frac{NS}{Q_1} = \frac{k}{\tau_a s + 1}, \quad \frac{NS}{v} = \frac{k^*}{\tau_a^* s + 1}, \quad \frac{NS}{Q_2} = 0
\]
The response of the heat affected zone width \( HZ \) to steps in either \( Q_1 \) or \( v \) can be closely approximated by a non-minimum phase second order behavior of the form:

\[
\frac{K_b(f_b^s+1)}{(r_1^s+1)(r_2^s+1)} = \frac{K_1}{r_1^s+1} \cdot \frac{K_2}{r_2^s+1}, \quad r_b = 0
\]

where the two modes can be associated with the dynamics of the two isotherms \( T_h \) and \( T_n \), the width difference of which defines \( HZ \). The sensitivity of \( HZ \) to the third input \( Q_2 \) is almost insignificant, as illustrated in Fig 12 by the very small steady-state deviations of \( HZ \) from the nominal conditions after a step in \( Q_2 \). For the completeness of the model a first order dependence of \( HZ \) on \( Q_2 \) will be considered here, so that:

\[
\begin{align*}
\frac{HZ}{Q_1} &= \frac{K_b(f_b^s+1)}{(r_1^s+1)(r_2^s+1)} \\
\frac{HZ}{v} &= \frac{K'_b(f'^b_b^s+1)}{(r'_1^s+1)(r'_2^s+1)} \\
\frac{HZ}{Q_2} &= \frac{K_b}{r_1^s+1}
\end{align*}
\]

Finally the response of the centerline cooling rate \( CR \) to steps in either \( Q_1 \) or \( v \) may be approximately described by an over-damped second order behavior owing to the existence of two dominating real poles. The dependence of \( CR \) on the third input \( Q_2 \) is also of second order, but with the one pole dominating over the other, so that it can be simplified to a first order transfer function:

\[
\begin{align*}
\frac{CR}{Q_1} &= \frac{K_c}{(r_1^s+1)(r_b^s+1)} \\
\frac{CR}{v} &= \frac{K'_c}{(r'^s_b+1)(r'_b^s+1)} \\
\frac{CR}{Q_2} &= \frac{K_c}{r_b^s+1}
\end{align*}
\]

The values of the gains of all the proposed transfer functions are easily determined from the steady-state values of the respective outputs (at \( t=0 \) and \( t=\infty \)) and the magnitudes of the steps imposed to the respective inputs. The time constants are calculated by a weighted averaging of those values that make the linearized responses to match (go through) the experimental data (or the simulated response in the case of \( Q_2 \) at the corresponding sequence of time instants. This is done separately for each of the two modes of the heat affected zone width, as well as for those of the cooling rate, where the secondary time constants \( r_b \) are selected to make the second order transfer functions match the initial flat part of the nonlinear response. The calculated values of the gains and time constants of the linearized model are compiled in Table 2. Note from this table the strong dependence of the parameters on the input magnitude.

Finally the linearized dependence of the welding outputs \( Y \) to the welding inputs \( U \) may be written in the form of a transfer matrix \( G \) as:

\[
Y(s) = G(s) U(s)
\]
where all the elemental transfer functions of $G(s)$ are as given above. As already mentioned, the responses of the linearized outputs above to the same step inputs are indicated on Figs 9-12 with a dot-dashed line.

CONCLUSIONS

Welding control systems must be considered a comprehensive system regulating all relevant weld properties in real-time. In this paper the process of GMAW is treated as a dynamically coupled, multivariable system in order to achieve optimal welding control system performance. A system description has been forwarded that defines the geometric and thermally activated properties of the weld in a fashion that is amenable to control system analysis. The basis for this analysis, multiple input-multiple output dynamic model, has been derived for both the case of detailed pool geometry and temperature field control. The character of both models is similar in that each is inherently non-linear, and simple linearization does not capture the true process operation and process disturbance effects.

Our current efforts are aimed at using both the "exact" and linearized, non-stationary models to develop control system strategies for multivariable weld control.

ACKNOWLEDGEMENTS

This work was supported by the US Department of Energy, under contract no. DE-FG02-85ER13331.

REFERENCES


---

**Figure 1. A Block Diagram Model of the GMAW Process**

**Figure 2. Schematic of GMAW with Definitions for the Geometry Control Model**
Simulated step responses of the non-linear weld bead geometry model for positive and negative inputs
Figure 9 Time responses of the outputs during the transient to the nominal conditions ($Q_1 = 2500W$, $v = 5mm/s$, $Q_2 = 0W$) step to the heat input $Q_1$ ($Q_1 = 2500W$ to 1.20 to 3000W).

Figure 10 Time responses of the outputs during a positive to the nominal conditions ($Q_1 = 2500W$, $v = 5mm/s$, $Q_2 = 0W$) step to the heat input $Q_1$ ($Q_1 = 2500W$ to 1.20 to 3000W).
Figure 11  Time responses of the outputs during a positive step to the torch velocity \( v (v=5\text{mm/s} \rightarrow 1.2v =6\text{mm/s}) \)

Figure 12  Time responses of the outputs during a positive step to the secondary heat input \( Q_2 (Q_2 =25\text{W} \rightarrow 0 \text{ source}=25\text{W}) \)
Table 1  Determined values of gains and time constants of the linearized weld bead geometry model

<table>
<thead>
<tr>
<th>Step</th>
<th>$K_1$</th>
<th>$K_2$</th>
<th>$K_3$</th>
<th>$K_4$</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$T_4$</th>
<th>$T_5$</th>
<th>$T_6$</th>
<th>$T_7$</th>
<th>$T_8$</th>
<th>$T_9$</th>
<th>$T_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>0.294</td>
<td>2.70</td>
<td>-0.367</td>
<td>-0.0831</td>
<td>-16.1</td>
<td>-0.613</td>
<td>0.218</td>
<td>-3.65</td>
<td>-41.7</td>
<td>818</td>
<td>2.33</td>
<td>-110</td>
<td>-0.199</td>
<td>41.7</td>
</tr>
<tr>
<td>$F_c$</td>
<td>1.0 in/s</td>
<td>0.294</td>
<td>3.15</td>
<td>-4.29</td>
<td>-0.0831</td>
<td>-16.1</td>
<td>-1.30</td>
<td>0.222</td>
<td>-4.20</td>
<td>-1.69</td>
<td>276</td>
<td>3.34</td>
<td>-110</td>
<td>-0.204</td>
</tr>
<tr>
<td>$F_c$</td>
<td>-1.0 in/s</td>
<td>0.294</td>
<td>2.20</td>
<td>-3.10</td>
<td>-0.0831</td>
<td>-16.1</td>
<td>4.69</td>
<td>0.212</td>
<td>-3.61</td>
<td>3.42</td>
<td>1.31</td>
<td>-130</td>
<td>-0.216</td>
<td>549</td>
</tr>
<tr>
<td>$F_d$</td>
<td>1.0 in/s</td>
<td>0.308</td>
<td>3.15</td>
<td>-4.51</td>
<td>-0.0663</td>
<td>-16.1</td>
<td>-1.30</td>
<td>0.222</td>
<td>-4.20</td>
<td>-1.69</td>
<td>276</td>
<td>3.33</td>
<td>-110</td>
<td>-0.204</td>
</tr>
<tr>
<td>$F_d$</td>
<td>-1.0 in/s</td>
<td>0.285</td>
<td>2.20</td>
<td>-2.95</td>
<td>-0.0569</td>
<td>-16.1</td>
<td>4.69</td>
<td>0.212</td>
<td>-3.61</td>
<td>3.42</td>
<td>1.30</td>
<td>-130</td>
<td>-0.216</td>
<td>4.69</td>
</tr>
<tr>
<td>$V_c$</td>
<td>0.1 in/s</td>
<td>0.294</td>
<td>2.70</td>
<td>-0.367</td>
<td>-1.104</td>
<td>-5.44</td>
<td>-0.613</td>
<td>0.218</td>
<td>-1.83</td>
<td>-41.7</td>
<td>0.10</td>
<td>1.18</td>
<td>-116</td>
<td>-0.218</td>
</tr>
<tr>
<td>$V_c$</td>
<td>-0.05 in/s</td>
<td>0.294</td>
<td>2.70</td>
<td>-0.367</td>
<td>-0.0488</td>
<td>-29.6</td>
<td>-0.613</td>
<td>0.218</td>
<td>-5.10</td>
<td>-41.7</td>
<td>0.90</td>
<td>3.20</td>
<td>-108</td>
<td>-0.202</td>
</tr>
</tbody>
</table>

Table 2  Determined values of gains and time constants of the proposed transfer functions for 1018 steel, 2mm thick, heat source Q' 10mm from primary source

<table>
<thead>
<tr>
<th>Output</th>
<th>Input</th>
<th>Nugget Section</th>
<th>Heat Affected Zone</th>
<th>Cooling Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Step</td>
<td>$X_a$</td>
<td>$T_a$</td>
<td>$X_b$</td>
</tr>
<tr>
<td>Q (W)</td>
<td>0.2X</td>
<td>2.45x10^{-3}</td>
<td>0.43</td>
<td>3x10^{-3}</td>
</tr>
<tr>
<td>V mm/s</td>
<td>0.2V</td>
<td>0.70</td>
<td>0.71</td>
<td>0.88</td>
</tr>
<tr>
<td>S</td>
<td>0.2</td>
<td>6.65</td>
<td>--</td>
<td>-0.088</td>
</tr>
<tr>
<td></td>
<td>-0.2</td>
<td>0</td>
<td>--</td>
<td>0.2</td>
</tr>
</tbody>
</table>
NEW DIRECTIONS IN NONIMAGING OPTICS

R. Winston, X. Ning, J.J. O’Gallagher, and W.T. Welford*

The Enrico Fermi Institute and the Department of Physics
The University of Chicago
Chicago, IL 60637

ABSTRACT

Current new directions in nonimaging optics are outlined. In ray optics, new three-dimensional solutions have been found which permit a transformation of the angular divergence of a beam of light without loss. A new approach to analyzing reversed optical systems is described. Such solutions and methods may be significant in fiber optic applications. In wave optics the radiance (or specific intensity) has been calculated for uniform lambertian sources. New effects are predicted which may be significant in diffraction limited measurements.

I. RAY OPTICS

1. New Solutions

A problem in geometrical optics which has been with us from the beginning of nonimaging optics is as follows: Given a rotationally symmetrical beam with uniformly filled phase space up to some divergence half angle $\theta_1$, how can we transform this beam, without intrinsic losses, to a beam with divergence half angle $\theta_2$ (Fig. 1). By intrinsic losses one means that rays are turned back. There are also optical losses due to absorption in refractive and reflective materials. These can be made extremely small by suitable choice and preparation of materials. By conservation of phase space the beam areas must change as $[\sin(\theta_2)/\sin(\theta_1)]^2$. The solution for $\theta_2 = \pi/2$ was suggested in a schematic way at our last Department of Energy Review three years ago and finalized in a recent publication of this group** (Fig. 2). It is based on the geometrical vector flux. The new solution is for $\theta_2 < \pi/2$ and can be called a $\theta_1/\theta_2$ transformer. It was found by Xiaohui Ning, a graduate student in our group. It is also based on the geometrical vector flux and bears the same relation to the previous solution as does the astronomical telescope to the Galilean telescope. Figure 3 shows how this design works and Figure 4 shows how well it works. Practical applications might be in launching light into multi-mode fibers as well as interconnecting optical fibers. It is particularly in the interconnect problem that the very low losses are of critical importance.

2. Reversibility in Geometrical Optics (This work is in collaboration with T. Jannson, Physical Optics Corporation, Torrance, CA 90505.)

a. Introduction

In the coupling of fiber optics and certain other applications one employs an

*W.T. Welford, Emeritus Professor of Physics at the Imperial College of Science and Technology, London, contributed to this work while a visiting professor at The University of Chicago.
Fig. 1. Lambertian over a restricted angle and the function of an optical transformer.

optical system coupled to its identical twin in reverse. This is shown schematically in Figure 5. It is usually desirable for the coupled system to have high throughput and it is tempting to assume that because geometrical optics is reversible, the identity of the two components of the coupler (C,C') automatically guarantees high throughput. This is not the case.

b. Reversibility in the phase space representation.

Let z be the optic axis of the system and introduce phase space coordinates \((x, y, z, P_x, P_y, P_z)\). Here the components of P denote the index of refraction of

Fig. 3. Transformer as the segment between two lens-hyperbola concentrators.

Fig. 4. Acceptance curves for different F numbers.
the medium times the direction cosine of a ray (the optical direction cosines). At the symmetry plane \( z = z_0 \) between \( C \) and \( C' \), the totality of light rays capable of traversing \( C \) and emerging there from (the phase space acceptance) will occupy some four dimensional volume in \( x, y, p_x, p_y \) bounded by a surface \( \Sigma \). To find what rays will also traverse \( C' \) we reason as follows: By the reversibility of geometrical optics (time reversal invariance) a ray with coordinates \( (x, y, z_0, -p_x, -p_y, -p_z) \) will certainly traverse \( C \). However, \( C' \) is the reflection of \( C \) in the \( z \) axis at the plane \( z_0 \). It follows that the ray with coordinates \( (x, y, z_0, -p_x, -p_y, p_z) \) will traverse \( C' \). We conclude that a ray will traverse the entire system consisting of \( C \) followed by \( C' \) if and only if both points \( (x, y, p_x, p_y) \) and \( (x, y, -p_x, -p_y) \) lie inside the volume bounded by \( \Sigma \).

Stating the condition for 100\% throughput in a different way, at every position \( x, y \) the surface \( \Sigma \) (which is now a curve in the \( p_x, p_y \) plane) must be symmetric with respect to inversion in the origin in momentum space.

The discussion simplifies for systems with axial symmetry about \( z \) which is the case of greatest practical interest. Because of this symmetry we can reduce the coordinates to: the radius \( r \), the radial component of momentum \( P_1 \) and the azimuthal component of momentum \( P_2 \). Then at each radius \( r \), the surface \( \Sigma \) may be visualized as a curve in the \( P_1, P_2 \) plane. Because of the azimuthal symmetry, the curve is symmetric in \( P_2 \) so that we need consider only the upper half plane. The condition for 100\% throughput becomes the requirement that at every radius \( r \) the curve \( \Sigma \) be symmetric with respect to reflection in the \( P_1 \) axis.

One way (but not the only way) this property can be secured is for the curve \( \Sigma \) to be a circle. Elements with this property have been called ideal. Very few examples of ideal axially symmetric optical elements have been found. These were described in Section I. However, nearly ideal elements can be generated by rotating ideal two-dimensional designs about the \( z \) axis. These are reflecting cones where the reflection is either specular or by total internal reflection. For example, a dielectric totally internally reflecting concentrator (transformer) is such a device.

c. Conclusion.

The throughput properties of an optical coupler consisting of an element joined to its symmetric twin are conveniently analyzed in terms of the phase space acceptance at the plane of symmetry. The requirement of high throughput imposes certain symmetry conditions on the phase space acceptance of the optical element. This description provides a method for calculating the throughput. More significantly, it provides insight into the design of couplers with very high throughput. Such systems are likely to be useful in the coupling of fiber optics.

II. WAVE OPTICS (GENERALIZED RADIANCE)

1. Introduction

During the last two decades several attempts have been made to formulate the foundations of radiometry. Numbers of definitions of radiance (specific intensity) have been proposed in terms of various kinds of second order field correlation functions. Although none of these definitions satisfies all the conditions that traditional radio-
metry requires\textsuperscript{6}, recent work indicates that most of them do agree with traditional radiometry in the small wavelength limit.\textsuperscript{7}

In this section, we investigate the generalized radiance function defined earlier\textsuperscript{2} for uniform lambertian sources in free space\textsuperscript{8} (uniform lambertian means the radiance is independent of position and direction). We will calculate the radiance functions for a few simple sources when the wavelength of the field produced by the source is small relative to the size of the source. The consistency of the generalized radiance with traditional radiometry at small wavelengths will be shown in general. Furthermore, some first order corrections to the generalized radiance function for strip shaded sources will be evaluated and their implications discussed.

2. Radiance Function for Uniform Lambertian Sources

Let us consider a source $\sigma$, occupying a finite portion of the plane $z=0$ and radiating into the half space $z>0$. We assume that the source fluctuations are statistically stationary. Furthermore, we assume that the source is uniform lambertian, which means that at the source plane, radiance is constant inside the source region and zero outside. For example, such a source might be realized physically by placing a plane aperture with dimensions much larger than the wavelength over the opening of a black body cavity. For simplicity, we consider unpolarized radiation at a specific wavelength.

A quantity crucial for our discussion is the second order angular correlation function $F(k,k')$ of the field produced by $\sigma$. Defined as an ensemble average of plane waves distinguished by wave vectors $k$ and $k'$, $F(k,k')$ can be written as follows:

\[
F(k,k') = \langle U^*(k) U(k') \rangle
\]  

where $U(k)$ is the angular component in the plane wave expansion of the field. If we use $V(\hat{r})$ for the field, then $U(k)$ and $V(\hat{r})$ are connected through a Fourier transformation.

\[
V(\hat{r}) = \int U(k) e^{i\hat{k} \cdot \hat{r}} d^3k
\]

$V(\hat{r})$ satisfied the Helmholtz equation $(\Delta + k^2)V(\hat{r}) = 0$ throughout the space $z>0$. For constant wavelength, all the $k$ and $k'$ above can be replaced by their corresponding unit directions $\hat{n}$ and $\hat{n}'$, and also equation (2) can be written as,

\[
V(\hat{r}) = (2\pi/\lambda)^2 \int U(\hat{n}) e^{i\hat{n} \cdot \hat{r}} d\Omega
\]

where $d\Omega = dLdM/N$ and $L,M$ and $N$ are direction cosines.

Using this notation, the complex generalized radiance function $B(\hat{r},\hat{n})$ can be defined in terms of $F(\hat{n},\hat{n}')$ as follows,

\[
B(\hat{r},\hat{n}) = F(\hat{n},\hat{n}') \left[ -i\hat{k} \cdot (\hat{n} \times \hat{n}') \right] d\Omega
\]

Let us consider the expression $\int B(x,y,z=0,\hat{n}) e^{ik[x(L-L')+y(M-M')]} dx dy$ over the entire $z=0$ plane. If one substitutes for $B(x,y,z=0,\hat{n})$ the radiance definition Eq.(4) and integrates over the $z=0$ plane first, this expression reduces to $\lambda^2 F(\hat{n},\hat{n}')/N'$. In other words, we can express the angular correlation function $F(\hat{n},\hat{n}')$ in terms of the generalized radiance function at the source plane $z=0$ as follows:

\[
F(\hat{n},\hat{n}') = N'/\lambda^2 \int B(x,y,z=0,\hat{n}) e^{ik[x(L-L')+y(M-M')]} dx dy
\]

Knowing the generalized radiance function at the source plane one can infer the second order angular correlation function of the field produced by the source. The
only assumption is that at the source plane, the generalized radiance function is the classical radiance. A uniform lambertian source of unit strength is represented by:

\[ B(x, y, z = 0) = 1 \quad (x, y) \in \mathcal{O} \]

\[ = 0 \quad (x, y) \text{ anywhere else.} \]

The angular correlation function \( F(\mathbf{n}, \mathbf{n}') \) can be written as

\[ F(\mathbf{n}, \mathbf{n}') = \int \int e^{ik\{x(L(L') + y(M(M'))\}} \, dx \, dy \quad \text{(over } \mathcal{O}). \]  

(7)

If we know the shape of the source, we can calculate the angular correlation. For a few simple cases, \( F(\mathbf{n}, \mathbf{n}') \) can be easily obtained:

\[ F(\mathbf{n}, \mathbf{n}') = N'\delta(L-L')\delta(M-M') \quad \text{for an infinite source;} \]

\[ F(\mathbf{n}, \mathbf{n}') = N'(2a/\lambda)\text{sinc}(ka(L-L'))\delta(M-M') \quad \text{for an infinite strip source with side } 2a; \]

\[ F(\mathbf{n}, \mathbf{n}') = N'(2a/\lambda)(2b/\lambda)\text{sinc}(ka(L-L'))\text{sinc}(kb(M-M')) \quad \text{for a rectangularly shaped source with sides } 2a \text{ and } 2b; \]

\[ F(\mathbf{n}, \mathbf{n}') = N'(2a/\lambda)^22J_1(ka(L-L')^2 + (M-M')^2)^{1/2}/ka(L(L')^2 + (M-M')^2)^{1/2} \quad \text{for a circular source with radius } a. \]

We observe that the cross-spectral density \( W(r_1, r_2) \) is related to the angular correlation function \( F(\mathbf{n}, \mathbf{n}') \) by an angular representation:

\[ W(r_1, r_2) = \int \int F(\mathbf{n}, \mathbf{n}') e^{i\mathbf{k} \cdot (\mathbf{n}_1 - \mathbf{n}_2)} \, d\mathbf{n}_1 \, d\mathbf{n}_2. \]

(8)

The generalized radiance function in the space \( z \geq 0 \) can be calculated using Eq. (4).

\[ B(x, y, z, \mathbf{n}) = \int f e^{ik[(x-x')(L(L') - L) + (y-y')(M(M') - M) - z(N(N') - N)]} \, dx' \, dy' \, dL' \, dM' \]

\[ (x', y') \in \mathcal{O}. \]

(9)

For the frequently encountered case that both the size of the source and the distance from the source are much larger than the wavelength, this integral can be evaluated explicitly. Under these conditions, the angular correlation functions are significant only when \( p = L(L') - L \) and \( q = M(M') - M \) are comparable to \( \lambda/a \) (a \( \approx \) physical size of the source).

Treating \( p \) and \( q \) as small variables, we can expand \( N-N' \):

\[ N-N' = (L-L')^2 - (L(L+L') - (M+M')^2)^{1/2} = \]

\[ (Lp+Mq)/(N+((L-L')^2 + (M-M')^2)^{1/2}) + \text{higher order terms}. \]

(10)

From Eqs. (9) and (10), the generalized radiance function \( B(x, \mathbf{n}) \) for lambertian sources can be calculated. If we only retain the first order terms in \( p \) and \( q \) (paraxial approximation) in Eq. (10), we reproduce the results of classical radiometry. Departures from classical radiance are expected if quadratic terms are also included in evaluating Eq. (9). By a suitable coordinate choice, we can set \( M=0 \) in Eq. (10).

Then,

\[ N-N' \approx -Lp/N + p^2/2N^3 + q^2/2N. \]

(11)
Substituting (13) into (8) and changing variables from $L'$ and $M'$ to $p$ and $q$ we obtain,

$$B(r, L, O, N) = \frac{1}{\lambda^2} \int \frac{i k ((x' - x + Lz/N)p + zp^2/2N^2 + (y' - y)q + zq^2/2N)}{dx'dy'dpdq}$$

(12)

Completing the square in $p$ and $q$, the integrations over $p$ and $q$ reduce to Fresnel integrals. Since $z \gg \lambda$, the $\lambda$ integrals converge rapidly and are very well approximated by Fresnel integrals over infinity. Then, equation (9) becomes,

$$B(r, L, O, N) = -i(\lambda^2/\lambda z) \int \frac{i k N^3(x-x'-Lz/N)^2/(2z) + ikN(y-y')^2/(2z)}{(x',y')d\sigma}$$

(13)

The following calculations are for a few simply shaped sources.

a. A strip source.

Let us assume the strip is infinite along the $y$ axis. Then the contribution from the $y$-component to the generalized radiance specified by equation (13) is a constant. The contribution from the $x$-component can be expressed in terms of Fresnel integrals $C(x)$ and $S(x)$ defined in the usual way:

$$C(x) = \int_0^x \cos(\pi t^2/2)dt \quad \text{and} \quad S(x) = \int_0^x \sin(\pi t^2/2)dt$$

(14)

Recalling that the real part of this function is identified with physical radiance, we can examine how closely this resembles the traditional radiometric descriptions of uniform lambertian sources, and also in what respect it departs from the classical picture.

Taking the real part, we have

$$\text{Re}(B(r, L, O, N)) = [C(u) + S(u) + 1]/2$$

(15)

where $u = N^3/(2\lambda z)^{1/2} (a-x+Lz/N)$.

In case the angle subtended by the source at the point of measurement is not too large, we may set $N \approx 1$ and express the dimensionless argument $u$ in terms of the angle $R = (a-x)/z + L/N$,

$$u \approx (2z/\lambda)^{1/2} \theta$$

(16)

A plot of $\text{Re}(B)$ vs $\theta$ for $(2z/\lambda)^{1/2} = 200$ is shown in Fig. 6. We notice several features which depart from traditional radiance.

1) $\text{Re}(B)$ is noticeably different from the classical radiometric picture when $\theta$ is comparable to $(\lambda/z)^{1/2}$. This suggests that the edge of the source is blurred over an angular width

$$\Delta \theta \sim (\lambda/z)^{1/2}$$

(17)

2) When this angular width is comparable to the angle subtended by the source, the traditional description should fail; that is, for

$$z \sim a^2/\lambda.$$

(18)

3) $\text{Re}(B)$ exhibits oscillations in the vicinity of the edge.

To discuss which of these features would be observable it is useful to consider an illustrative measurement at $\lambda \approx 10 \mu m$, $z = 20 \text{ cm}$ with a diffraction limited radiometer consisting of a lens focussed to infinity onto a detector. The angular
resolution obtainable with this apparatus is determined by the radius \( r \) of the lens
and its point spread function. The resolution can be visualized by projecting the
angular acceptance of the radiometer back to the plane of the source. This defines
an area, essentially a circle of radius \( R \approx r + 1.22 \lambda z/2r \), graphically referred to as
the "foot print" of the radiometer acceptance. Since this radius minimizes at
\( R = (0.61 \lambda z)^{1/2} \), the minimum obtainable angular resolution is \( R/z = (0.61 \lambda z)^{1/2} \)
which is comparable to the scale of the edge structure. One therefore expects, and
computer simulation confirms, that the oscillations average out. We also note that
averaging over a range of wavelengths, as would be the case for a realistic experiment,
also tends to wash out the oscillations. However, the rounding of the edge persists.
We conclude that the width of the edge should be distinguishable from the classically
sharp edge by a careful measurement. The results would be a critical test of the
foundations of radiometry.

3. A Rectangular Source with Sides a and b.

For this case, it can be shown that the generalized radiance function is also
complex. However, the real part of this function corresponding to the physical
radiance is the following:

\[
\text{Re}(B(0,0,z,L,0,N)) = (1/2) \int \int \sin(\pi u^2/2) \cos(\pi v^2/2) du dv + \int \int \cos(\pi u^2/2) \sin(\pi v^2/2) dv \]

where \( u = N((x-x') + Lz/N)(2N/\lambda z)^{1/2} \) and \( v = (y-y')(2N/\lambda z)^{1/2} \).

Since \( a \) and \( b \) are much larger than \( \lambda \) the Fresnel integrals \( \int \cos(\pi v^2/2) dv \) and
\( \int \sin(\pi v^2/2) dv \) in Eq. 19 can be replaced by 1.

It is apparent that the mathematical formalism is identical to that of the strip
source. We point out that even although the 1st order corrections to the generalized
radiance have been evaluated only for strip and rectangular sources, the two features
we have observed would persist for any source.
4. Conclusions

We have shown that the 0th order approximation of the generalized radiance functions coincides with the traditional radiometric description for lambertian sources. However, within an angular spread $\approx (\lambda/z)^{1/2}$ pointing back to the source edge, or at a distance comparable to $a^2/\lambda$ from the source plane, the generalized radiance functions do not agree with traditional radiometry. The angular spreading of the edge has a different wavelength dependence than diffraction in a measuring instrument and the effects should therefore be distinguishable. It would be interesting to test this experimentally.

III. ACKNOWLEDGEMENT

This work was supported by Department of Energy grant number DE FG02 87ER 13726.

IV. REFERENCES


7. J.T. FOLEY and E. WOLF, Optics Comm.55, 236 (1985); this conclusion is actually proved for one particular definition of the radiance function, essentially the one used here. However, the paper suggests that the conclusion holds more generally.


WKB EVOLUTION OF WAVE PACKETS

Robert B. Littlejohn
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720 U.S.A.

ABSTRACT

Wave packet methods constitute a new approach to WKB theory, and provide both a new theoretical tool and a new way of solving practical problems. They are especially useful for complex geometries in several dimensions. Their main advantage is the ease with which they deal with caustics and Maslov phase shifts, and the fact that they do not suffer from any discontinuities or divergences.

INTRODUCTION

Wave packet methods and other methods closely related to them are emerging as an interesting alternative to traditional WKB theory for certain applications involving the asymptotic properties of wave fields. Especially in multidimensional problems with complex geometry or those with a complicated caustic structure, wave packets offer certain advantages over WKB theory. In other kinds of problems, such as those with only one dimension of spatial variation, WKB theory is usually easier to use. In still others there may be advantages to a hybrid approach. In any case, wave packets bring a new perspective to the whole subject of wave asymptotics, and provide new theoretical tools which can be of great use.

Whereas WKB theory is based on a certain ansatz for the wave field, i.e. one involving a rapidly varying phase and a slowly varying amplitude, the approximation inherent in wave packet methods is based on the fact that wave packets, in view of the Heisenberg uncertainty principle, represent objects which are localized in the x-k phase space. That is, when we think of a wave packet and its Fourier transform as two manifestations of the same object, we are naturally led to combine the pictures and imagine a single object in the ray phase space, whose dimensions satisfy \( \Delta x \Delta k = 1 \). In a picture which has been developed by Heller, Markevitz and myself, it turns out that a wave packet can be associated with a localized ensemble of "ray particles" in the ray phase space. These particles evolve in time according to Hamilton's equations, based on the ray Hamiltonian, which is just the frequency or local dispersion relation, \( \omega = \omega(x,k) \). As a result, the center or expectation value of the wave packet obeys the Ehrenfest relations, i.e. it follows a single ray trajectory.

The notion of a wave packet (or other wave field) as an object in the ray phase space can be made precise by the use of the Wigner function. For example, the Wigner function of a wave packet is a function \( w(x,k) \) which looks very much like
a localized probability density function, which, for the sake of vividness, can be thought of as a distribution of "wave particles." The Wigner function is defined by

\[ W(x,k) = \int d^{3}s \ e^{-ik \cdot \phi(x + \frac{s}{2}) \ \phi(x - \frac{s}{2})}, \]  

where \( \phi(x) \) is the wave field in question.

In principle the Wigner function can be used to perform exact calculations of wave phenomena, with no approximation. In the event, however, that the Wigner function is localized, as is the case with the Wigner function of a wave packet, it lends itself to a simple approximation which contains the essence of the approximate wave packet evolution as well. In this approximation, the center of the Wigner function follows a ray trajectory, as in the Ehrenfest relations, and the particles on the wings of the Wigner distribution are evolved by the linearized ray dynamics about the central ray. Thus, the approximate evolution of the localized Wigner function consists of a translation of the center along the ray trajectory, and a "rotation" about the center, as described by the linearized behavior of nearby trajectories. I put the word "rotation" in quotes because the linearized behavior of Hamiltonian trajectories does not give rise to orthogonal rotations in the usual sense, but rather to the so-called symplectic transformations. These include stretching operations as well as ordinary rotations, and are characterized by the fact that they preserve the x-k area in phase space. They are well known in stability studies of mechanical systems.

Thus, a simple approximation is available for the evolution of localized Wigner functions. But how does this transcribe into the evolution of actual wave packets? The answer is by the use of displacement and rotation operators, which form so-called projective representations of the corresponding operations in the ray phase space. The concept of projective representations is well known in certain areas of mathematical physics, but what it comes down to in this instance is that the evolution of wave packets carries phase information which is not present in the evolution of localized ensembles of particles.

This phase can be broken into two parts. One is essentially the Bohr-Sommerfeld phase, which is familiar from traditional WKB theory. This phase is, of course, well known, although wave packet methods provide an interesting new perspective on it.

The other part is more interesting. It is the phase connected with the Maslov indices of the WKB wave function. In WKB theory for one-dimensional systems, this phase appears as a discontinuous jump of \( \pi/2 \) at each turning point, where the wave function itself diverges. In wave packet theory, however, this phase accumulates smoothly and continuously along a ray trajectory, and there are no divergences or discontinuities of any sort in the wave packet itself. It is in this respect that the advantage of wave packet methods over WKB theory is strongest: wave packets are completely oblivious of caustics or turning points, and require no special attention on account of them. This advantage is especially strong in higher dimensions, where the existing WKB theory for dealing with caustics is very complicated, and where the number of caustic types proliferates.

Since the multidimensional theory of the Maslov index in the WKB picture is complicated, and since wave packets handle the Maslov index and caustics so easily, one might suppose that wave packets could be used as a theoretical tool for gaining deeper insight into the Maslov index. Indeed, this is the case, as shown by some recent work by J. Robbins and myself, in which we use wave packet theory to find new ways of computing the Maslov index. Although the Maslov index is usually conceived of as a caustic count, the new formula does not involve caustics; instead, it is based on topological properties of certain spaces (the symplectic group manifold and its homogeneous spaces). We have applied our formula to computing the Maslov indices of resonant tori, a problem which is important in the quantizing of nonlinear systems in atomic and molecular physics.

Problems involving spatially extended wave fields, as most do, can be handled by wave packet methods by forming linear combinations of wave packets. Such representations are considerably more efficient than the usual orthogonal
expansions, which involve basis sets such as plane waves or Hermite functions. I will report in the future on some studies of normal modes of multidimensional wave fields (i.e. eigenfunctions), which are represented in terms of wave packets.\(^6\)

Wave packets are a new tool, and they are not as extensively studied as WKB theory. Therefore these are certain areas in which wave packet theory is not well developed, such as in tunnelling processes. It seems likely, however, that wave packet methods could be applied here, as well, and would lead to theoretical insights. The whole subject of multidimensional tunnelling for nonseparable systems is poorly developed, and could use the benefit of a new approach. Another such area is that of diffraction processes, which is important in many applications. Although simple calculations indicate that wave packets could be used for such applications, it seems that little work has been done on this. Altogether, wave packets have made an important contribution to the theory of wave asymptotics, and show promise of continuing to do so.

ACKNOWLEDGMENTS

This work was supported by the Department of Energy under contract number DE-AC0376SP00098, and by the National Science Foundation under Grant No. NSF-PYI-84-51276.

REFERENCES

Chaotic Advection: The Efficient Stirring of Viscous Liquids

Hassan Aref and Scott W. Jones
University of California, San Diego
La Jolla, CA 92093, USA

Abstract
The theory of dynamical systems provides a useful framework for discussing the motion of passively advected particles by a flow. Particle orbits can be either regular or chaotic. The different regions of the flow, where one or the other of these possibilities is realized, can be identified by the construction of Poincaré sections and the computation of Lyapunov exponents. Efficient stirring comes about by arranging flow control parameters such that many particle trajectories display large scale chaos. The kinematic origin of chaotic advection allows it to occur even in flows at very low Reynolds number. Results for three flow situations are discussed: the Stokes flow between eccentric, rotating cylinders; the advection by a pulsed source/sink system; and laminar flow in a twisted pipe.

INTRODUCTION
For convenience we shall follow a terminology suggested by Eckart in which stirring signifies the mechanical process whereby fluids are distributed more uniformly within a given domain, i.e. stirring is a process of stretching of intermaterial area. Mixing, on the other hand, is the process of diffusion of substances across intermaterial surfaces. Stirring can promote mixing by creating more intermaterial surface area. Mixing depends on material properties, such as diffusivities, whereas stirring is a purely kinematical aspect dependent only on flow parameters. Indeed, it is possible to stir fluids that do not mix at all.

Fluids are stirred and mixed routinely in Nature and in a wide variety of technological applications. However, our understanding of what controls the efficiency of stirring is not in an altogether satisfactory state. Furthermore, the "language" that has traditionally been used by chemical engineers and others for the discussion of stirring processes has until recently been unnecessarily limited. The main thrust of our work is to show how application of the theory of dynamical systems to the kinematics of stirring can lead to a deeper understanding, and to illustrate this by the analysis and numerical simulation of sample flow situations. With a better understanding of what makes a given stirring process more efficient we can enhance our ability to predict such things as the dispersal of pollutants, and to improve the design of commercial "mixers" in a variety of situations. One of the prime results of our efforts has been the elucidation of how to achieve efficient stirring by laminar flows, a problem that arises in the processing of very viscous liquids.

Much stirring in practice comes about by turbulent flows. It is intuitively clear that one must expect to see stirring and mixing on all the scales of such a turbulent flow. At the other extreme a substance can fill up a completely stagnant volume of fluid simply by diffusion due to molecular collisions between molecules of the substance and those of the fluid. Diffusion is typically a slow process. Sometimes one substance will have a very low rate of diffusion into another.
In many applications turbulent stirring is not an option. The cost of homogenizing very viscous liquids as occurs, for example, in the processing of polymers is prohibitive. Turbulence implies high stresses within the fluid that may be undesirable. For example, in the processing of biofluids turbulent stirring can damage the long chain molecules that give the fluid its biological characteristics.

So how does one stir efficiently if turbulence is ruled out for one reason or another, and diffusion is so slow that one is not willing to wait for it? One answer is "by chaos," and numerical computer "experiments" on this kind of stirring are a prime objective of the work reported here.

Chaos (in the technical sense) refers to a phenomenon that has recently become much studied in several branches of science. In mathematical terms chaos or chaotic behavior is the notion that very simple equations can have extremely complex solutions. Indeed, even though the equations can be written down using only elementary concepts of algebra and calculus, the solutions show intricate patterns of complexity, with a hierarchy of scales and with such a sensitive dependence on initial conditions that often only a probabilistic characterization is possible. This scenario is well known in a number of guises.

For passive advection of fluid particles by a flow field \( \mathbf{V}(x,y,z,t) = (u(x,y,z,t), v(x,y,z,t), w(x,y,z,t)) \) the equations of motion are

\[
\frac{dx}{dt} = u(x,y,z,t); \quad \frac{dy}{dt} = v(x,y,z,t); \quad \frac{dz}{dt} = w(x,y,z,t)
\] (1)

Clearly these are in general rich enough to admit chaotic solutions. Indeed, in three dimensions this can be achieved for steady flow as we shall see when we study stirring of a fluid in a bent and twisted pipe.

In two dimensions the dependence on time \( t \) is required. For incompressible flow the development can be taken one step further: If we introduce the streamfunction \( \psi(x,y,t) \), the velocity components are spatial derivatives of \( \psi \) and in place of (1) we have

\[
\frac{dx}{dt} = \frac{\partial \psi}{\partial y}; \quad \frac{dy}{dt} = \frac{\partial \psi}{\partial x}
\] (2)

Note that Eqs.(2) have the form of Hamilton's canonical equations for a one degree of freedom system and that \( \psi \) plays the role of the Hamiltonian. Several important points are worth making: (i) the system (2) is integrable if it is autonomous. This corresponds to the case of steady flow, where it is well known that particle paths follow streamlines. (ii) The Hamiltonian nature of the dynamics (2) is valid even if the flow is dominated by viscous forces. It is a manifestation of the kinematic constraint of incompressibility, not of any particular aspect of the momentum equation governing the fluid motion. (iii) In this Lagrangian representation of the problem of advection the introduction of dynamical systems theory requires no arbitrary truncation to a small set of modes as is often the case when using the conventional Eulerian representation of flow quantities.

The idea behind stirring by chaotic advection is to have a (simple) flow field arranged such that many of the particles being advected run along very complex (chaotic) trajectories [1]. If this can be set up, efficient stirring will take place. Note that this is done without appealing to any intrinsic complexity of the flow itself, and without waiting for diffusion. That this is possible and practical was first shown in computer experiments where the motion during the stirring process was simulated [1,2].

210
Laboratory verifications have followed [3,4]. In their simplest form the calculations are not very demanding and can be done on a desktop computer. On a supercomputer, such as the CRAY-XMP at the San Diego Supercomputer Center which we have used, it is possible to follow up to 100,000 stirred fluid particles with relative ease. This is important in order to pursue fine details of the stirring process.

We have used the term “efficiency” in a rather loose way. This reflects the fact that any quantitative measure of efficiency that one chooses will show an enhancement of stirring quality as more and more of the particle orbits become chaotic. In particular, it is important to realize that the existence of chaotic advection implies a complementary regime of integrable advection. In this regime the coordinates of particle orbits are quasiperiodic functions of time. Small perturbations of the integrable regime are dominated by stability results of the KAM type. Hence, chaotic advection can also be viewed as the converse of regular or integrable advection regimes where impenetrable barriers exist to transport.

The transition from integrable to chaotic advection is intrinsically different from the transition from laminar to turbulent advection. The latter typically involves enhanced agitation of the fluid. To create the turbulent flow one needs to stir more vigorously. In general this means also spending more energy for a given degree of homogenization. Achieving chaotic advection typically implies no additional energy consumption. In 2D we simply substitute a pulsed flow for a corresponding steady flow. In our 3D pipe flow example the only change is in the geometry of the bent pipe: A bent pipe and a bent and twisted pipe transporting fluid at the same pressure gradient can have profoundly different stirring characteristics, as we shall elaborate below.

![Figure 1: Sample particle trajectories in a pulsed source/sink system (after [8])](image)
CHAOTIC ADECTION IN A STOKES FLOW

The independence of the phenomenon of chaotic advection on usual flow parameters measuring agitation, in particular Reynolds number, has as its corollary the interesting and important fact that efficient stirring via chaos can be accomplished for creeping flow. A familiar and useful example to keep in mind is the kneading of dough. The so-called "Baker's transformation" was inspired by this, and is, of course, a paradigm of chaotic motion [5]. We have in mind cases in which a Stokes flow stretches and folds a very viscous, two-component liquid producing ever finer striations. An interesting example where this can be followed in detail is the flow between eccentric rotating cylinders. Numerical and laboratory experiments on this flow have been reported in the literature [2,4,6] showing remarkable agreement and completely verifying the notion that chaotic advection patterns can indeed arise even in laminar flow. The (incorrect) perception that a passive scalar must mimic the smoothness of the flow that stirs it still dominates much conventional thinking on this problem. It arises from the Eulerian version of Eqs.(1), viz

\frac{\partial \theta}{\partial t} + \mathbf{v} \cdot \nabla \theta = 0

(3)

where \( \theta \) is an advected scalar field, and the facile statement that (3) is a linear equation in \( \theta \). The realizations that advection by Stokes flow can lead to efficient stirring via chaotic particle motions, and that several aspects of the theory of dynamical systems provide the appropriate terminology for discussing the control of this kind of stirring, are extremely important for a variety of applications.

BLINKING VORTEX, PULSED SOURCE/SINK AND OTHER MODEL FLOWS

As in the general theory of dynamical systems a few simple model flows have been studied because of their value in providing insight. The blinking vortex flow was introduced in [1] and has since been studied further in considerable detail [7]. In this model one studies the stirring produced by two fixed point vortices that are switched on and off alternately. If the flow plane is parametrized as the complex z-plane, this flow gives rise to a mapping \( z \rightarrow z' \):

\begin{align}
z' &= -1 + (1-z) \exp \left( \frac{i \mu}{|1-z|^2} \right) \\
\mu &= \frac{\Gamma T}{4\pi a^2}
\end{align}

(4a)

(4b)

is the basic control parameter made up of the circulation of the vortices, \( \Gamma \), the switching period, \( T \), and the distance between the vortices, \( a \). In (4a) the particle position iterates are measured in units of \( a \).

An intriguing generalization of the blinking vortex flow, Eqs.(4), is the pulsed
Figure 2: Sensitive dependence on initial conditions in the large: Two circular blobs of fluid are started close together in the pulsed source/sink system. One blob disperses completely, the other is confined by KAM tori and remains intact. (After [8,9]).

source/sink flow, studied recently by the present authors [8]. In this case we have a potential source and sink separated by some distance. The source and sink are turned on alternately. Particle trajectories now are polygon lines made up of line segments aligned (alternately) with the radial directions from the source and sink. Examples are shown in Fig.1. One observes that each particle goes down the sink after a finite number of strokes. Hence, to use this setup for stirring fluid, it is necessary to recycle the fluid from the sink to the source. This may be done in different ways. If we envision collecting the fluid taken out by the sink during one stroke in a "cylinder" beneath the flow plane, there are clearly infinitely many options for reintroducing that amount of fluid into the flow on the next stroke of the source. We have discussed three typical methods in our paper [8], one of which has the interesting feature of leading to integrable advection which can then be perturbed systematically.

Here we show just one other result from this work, that we refer to as sensitive dependence on initial conditions "in the large." In Fig.2 the fate of two blobs of marked fluid is shown. One of them stays together even after repeated reinjection cycles through the sink/source system. The other is dispersed by chaotic advection. This figure clearly shows the qualitative differences of these two advection regimes. We note also that the "islands" seen in Fig.2(d) are not due to vortical motion of the fluid, because the pulsed source/sink flow field is a potential flow at every instant in time. The barriers to transport, so explicit in Fig.2, arise from the KAM theorem applied to a time-dependent potential flow (with particle reinjection), and reflect the phase
mismatch between particle motion through the system on different cycles of the reinjection. The parametric study in [8] shows how to control the number and size of such islands. This example was also used to highlight the sensitivity of Lagrangian data in numerical flow simulation techniques [9].

CHAOTIC ADECTION BY LAMINAR FLOW IN A TWISTED PIPE

A flow system of considerable interest in a variety of applications both in mechanical and bioengineering is the flow in a curved pipe. It has been known for a long time that inertia produces transverse, secondary flows in such a pipe in the form of counterrotating, longitudinal vortices. Perturbation solutions for the flow were derived already by Dean [10] some 60 years ago.

The secondary vortices have an axis of reflection in the plane of curvature of the pipe. Thus, if a piece of curved pipe is followed by another with a plane of curvature rotated through an angle relative to it, particles travelling along this twisted pipe will experience secondary vortices with one orientation followed by similar vortices turned through some angle in the cross-sectional plane. A long sequence of such curved pipe sections, each one turned with respect to its predecessor should then have a similar effect on transverse transport as a succession of "blinking vortex pairs" of different orientations. A recent investigation [11] is aimed at corroborating this idea and providing quantitative data on the degree of stirring achievable as a function of the angle between successive curved pipe sections and the rate of flow.

The basic unit of the pipes we consider is a 180° curved segment of sufficiently large radius of curvature relative to its radius that the approximations of Dean and others [10] are valid. We follow the 180° segment by another of similar geometry but rotated through some angle $\chi$ relative to it. If $\chi=0^\circ$, the resulting pipe is a torus. If $\chi=180^\circ$, the resulting pipe is plane and S-shaped. For $\chi>0^\circ$ we envision following this two-segment "basic cell" by a large number of identical replicas. We have also considered twisted pipes made up of repeat units with four segments.

Figure 3: Poincaré sections for transverse advection in a twisted pipe. The panels trace the change in distribution between regular and chaotic regions as the pitch angle $\chi$ is varied at a fixed flow rate.

(a) $\chi=\pi/4$; (b) $\pi/2$; (c) $3\pi/4$. (After [11]).
The flow through this pipe is modelled by "glueing together" the appropriate Dean solutions, turned as required at each junction. We study the mapping of particle positions in a circular cross-section that sends a particle to its position a basic cell length (along the pipe) later. This is not an immediately accessible quantity, since real particles advecting down the pipe will not arrive at cross-sections spaced by a basic cell at the same time because the axial flow is not uniform over the cross-section. Nevertheless, it is a useful quantity because it picks out the cross-sectional imprint of tubes of flow that wind down through the entire pipe, and, by contrast, the regions of chaotic, transverse advection. Figures 3 and 4 show examples of the iterates of this mapping for different values of the flow rate and for changes in the angle of pitch $\chi$.

![Figure 4: Poincaré sections for transverse advection in a twisted pipe. The panels trace the change in distribution between regular and chaotic regions as the flow rate is changed for fixed value of the pitch angle $\chi$. Flow rate decreases from (a) to (c). Fig.3(b) is intermediate between (a) and (b). (After [11]).](image)

CONCLUDING REMARKS

It is clear that there are numerous instances in which the concept of chaotic advection is useful and important. We envision developments in the following areas:

(i) **The relation to turbulent mixing**: The so-called viscous-convective subrange [12] in isotropic, turbulent mixing of a passive advectant constitutes a situation where the scales of the advectant are substantially smaller than the smallest scales of the flow field. One must expect, in general, that these smaller scales arise by chaotic advection, and one can ask whether the theory of dynamical systems can predict a power law spectrum for the advectant autocorrelation. Another interesting set of problems of considerable practical importance arise from the interplay between chaotic advection on large scales and turbulent mixing of small scales of a given flow.
(ii) **Corrections to the passive advection model:** In general entirely passive advection as described by Eqs.(1-3) is too idealized to provide a complete description. A multitude of additional effects must be added. For fluids the diffusion of one into the other is usually not an entirely negligible effect, especially not in situations where the intermaterial area is being increased significantly. Diffusion can be modeled in the Lagrangian representation by turning (1) into a Langevin equation. The boundary between one fluid and another in the near immiscible limit is usually endowed with surface tension, and the densities and viscosities of the two fluids being stirred are not always exactly matched. We have previously introduced the terminology of active and passive interfaces in this context [13] and shown that many other aspects of a deforming interface must now be taken into account. For the advection of discrete particles flow drag and particle inertia have a role to play.

(iii) **Extensions to "open flows":** The pulsed source/sink system gives us a first hint of situations where chaotic advection may be useful but where we are not dealing with a closed system. Other situations of this type are residence time distributions and entrainment problems. The basic scenario is to have particles that interact with a basic flow for a finite time, but to have that time be a sensitive function of initial data. In this way these problems resemble scattering problems. Chaotic behavior in scattering problems has recently become a vigorous object of research since so many of the traditional concepts from closed systems, such as Poincaré sections and Lyapunov exponents do not immediately carry over.

(iv) **Applications to flow visualization:** The tracking of dye, smoke, bubbles and other essentially passive advectants has become a favorite method of visualizing flow features in the investigation of turbulent shear flows. It is clear from the concept of chaotic advection that such signals can have considerably amplified levels of "noise" in the sense that the particle tracks can be very complex even though the underlying flow is really quite simple. For example, a "bursting" event as seen in shear flows just before the transition to "fully developed" turbulence may be less of a flow instability and more of a transition to chaotic advection regimes than has been appreciated thus far. The concept of chaotic advection certainly suggests a sober reexamination of several results based on flow visualization, and amplifies the conclusion that coherent flow visualization signals imply coherent advecting flows.

**ACKNOWLEDGMENTS**

The work reported in Refs.2,8,9 and 10 is supported by DoE, Basic Energy Sciences, under grant DE-FG03-85-ER13. Supercomputing resources made available at the San Diego Supercomputer Center were supported by NSF/PYI grant MSM84-51107. We are indebted to Oran M. Thomas for writing several interactive graphics programs for use on our Iris workstation. These efforts were supported in part by DARPA/ACMP under URI grant N00014-86-K-0758 administered by ONR.
REFERENCES

MIXING: STRETCHING, BREAKUP, AND CHAOS

J. M. Ottino
University of Massachusetts
Chemical Engineering Department
Goessmann Laboratory
Amherst, MA 01003-0011

ABSTRACT

In spite of its universality and practical implications, the understanding of fluid mixing remains largely empirical and little theoretical foundation exists at the present time. In this paper we summarize, briefly, some of our recent work in this area. Two closely related matters are discussed: (i) theoretical and experimental investigations of the conditions that determine the onset of chaos in deterministic two dimensional flows, and (ii) studies of the dynamics of stretching and breakup of small droplets in such flows. Thus, (i) focuses on a global scale whereas (ii) is a local analysis focusing on deformation and breakup in Lagrangian (linear) time dependent flows. The breakup depends in a complicated way on the balance between elongation and vorticity at small scales; on the other hand, the dispersion is effective if the large scale flow is chaotic. Much work remains to mesh the two levels; in this paper we elaborate on both aspects and indicate possible lines of attack for future studies.

INTRODUCTION

Usually, the starting point in a mixing process is fluids segregated on a large scale consisting of blobs or striations. Mechanical mixing causes the striations to stretch and, depending on interfacial tension, viscosity differences, and so on, to break into droplets or ribbons. The stretching and breakup are related in a complicated way to the velocity field. In most cases of interest it is impossible to take into account all the aspects of the problem (there might be coalescence too) and the objective of a theoretical framework is to produce a simple picture that captures the essential physics of the process. Such a framework can be useful in the understanding of problems as diverse as mixing in the Earth's mantle [1], mixing in oceans
[2,3], and the problem that originated this work, mixing of viscous liquids such as polymers [4].

A major simplification of the analysis is possible if the problem is decoupled into global and local components. In the first part of this paper we focus on the global aspects, which correspond to the mixing of a single fluid. Here, the emphasis is on the relative contributions of the regular and chaotic regions produced in two-dimensional chaotic flows. In the second part, we address the local aspects by considering the effects of the macroscopic flow on the behavior of microstructures characterized by some evolution equation. The picture emerging from these two contributions is far from complete and more work is necessary.

GLOBAL ASPECTS: EXPERIMENTAL AND THEORETICAL RESULTS

A program for an investigation of mixing in two-dimensional flows should focus on the equations for the fluid particle trajectories in the x-y plane, i.e.,

$$\frac{dx}{dt} = \frac{\partial \psi}{\partial y}, \quad \frac{dy}{dt} = -\frac{\partial \psi}{\partial x}, \quad (1)$$

and try to isolate phenomena of a generic nature; i.e., results relevant to a wide class of flows and not just on the details of particular flows. Fortunately some theoretical guidance exists for this case: the system is Hamiltonian with one degree of freedom if the flow is steady, and with two if the streamfunction is time dependent [5-6].

Much understanding can be gained by detailed analysis of prototypical configurations. Investigations can proceed in experimental as well as theoretical/computational directions and both aspects should be regarded as complementary (more on this later on). On the experimental side we have examined two time-periodic flows: a cavity flow [7] and a journal bearing flow. Both experimental systems are versatile and both have advantages and disadvantages. For example, compared to the journal bearing flow the cavity flow mixes better and less periods are necessary to achieve significant effects but, on the other hand, it is more susceptible to experimental error, and there is no analytical solution for the streamfunction. A preliminary set of results for the cavity flow was reported by Chien, Rising, and Ottino [7] and a more complete set of results focusing on the comparison between computational and experimental studies is in preparation. Since most of the work on the journal bearing flow is in progress and space is limited, we will focus on the periodically operated cavity flow. The apparatus is capable of producing a two-dimensional creeping flow in the x-y plane (for details see [7]). The flow region is rectangular with width $d$ and height $H$ (other
configurations are possible). Consider, as a special case, that the top wall (denoted 1) and the lower wall (denoted 2) are moved via computer control with velocities $v_1$ and $v_2$ of the form

$$v_1 = U_1 \sin^2(2\pi t/T_1)$$

$$v_2 = -U_2 \sin^2(2\pi t/T_2 + \alpha)$$

with equal amplitudes, $U_1=U_2$ equal time periods, $T_1=T_2=T$ and a phase angle, $\alpha=\pi/2$. The prescription (2.1-2) acts as a forcing condition of the system given by equations (1), via a boundary condition on the Navier-Stokes equations (if the fluid is Newtonian), and produces a time-periodic velocity field. Obviously, there are infinitely many other possibilities for the motion of the walls. For example, even within the case (2.1-2), we can change the relative amplitudes, the phase angle, and the frequencies. In the results presented here $T$ is the governing parameter.

If the flow is slow (both the Reynolds and the Strouhal numbers are small), the instantaneous streamlines form closed orbits with no secondary vortices and the elliptic point at $x=l/2W$ moves up and down in a time periodic fashion in the $y$-interval $1/3H-2/3H$. It is known that the cavity flow is capable of producing Smale horseshoe maps [7]. A consequence of horseshoe formation is a mixture of regular and chaotic regions which are clearly seen in experimental studies (Figure 1). A knowledge of the behavior of the elliptic islands and horseshoe maps is valuable in deciphering the behavior of the flow upon perturbations of the forcing parameter, $T$. Such a knowledge is rather limited at this time.

Even though the experiments of Figure 1 give rise to smooth folds and striations, the difficulties in obtaining a similar picture by conventional computational means [8] presents nearly insurmountable problems in resolving thin striations and folds. For example, we estimate that in order to resolve a length stretch of order $10^4$ we need to track at least $10^8$ points. Such a computation on a one mega-flop machine could take approximately 3 years and use 600 Gbytes of memory (detailed calculations will be reported at a later date). This is a powerful reason for the marriage between computations and experiments in the analysis of chaotic mixing.

The theoretical studies can focus on aspects that are hard to access by means of experiments. For example, we have developed methods to locate period-1 periodic points in both flows and computed the stable and unstable manifolds associated with the hyperbolic points in the journal bearing flow (if the location of the periodic points can be established it is then in theory possible to generate the unstable manifolds experimentally). The level of complexity of the analytical studies for systems which can be studied in the laboratory--such as the journal bearing flow and the cavity flow--is considerably more difficult than those of mappings (Khakhar, Rising, and Ottino [6]). Fortunately, the treatment of horseshoes in discrete maps and
continuous systems is similar and they can be explored both analytically and computationally.

Based on experimental and computational studies to date, it appears that the structure of the system is largely dominated by low period Smale horseshoe maps (the period of the horseshoe being the smallest fundamental period of the invariant set) and the presence of elliptic islands (the period here is the period of the central elliptic point). However, many of the islands that appear in our experimental studies (see Figure 1) do not appear to conform to the typical picture of an elliptic island and appear to be the result of what might be termed incomplete horseshoes. A thorough experimental documentation of this fact is needed.

From the point of view of the microstructures what matters is the linearized flow around a material particle. The stretching history in the regular regions leads to linear length increase; the stretching over many periods in the chaotic regions is exponential. We thus expect that the behavior of microstructures in both regions will be substantially different. However, much more work is necessary to characterize the stretching in both regions. Such histories enter as an input into the equations governing the stretching and breakup of microstructures in chaotic flows.

**LOCAL ASPECTS: DEFORMATION OF DROPLETS AND MICROSTRUCTURES IN UNSTEADY FLOWS**

The simplest problem of stretching of microstructures corresponds to the case of passive material elements. In the case of passive mixing, which can be regarded as a case of two fluids of similar viscosities and without interfacial tension, the boundary between the two fluids acts as a marker of the flow; the motion is topological (i.e., there is no breakup). In the case of active interfaces the interfaces interact with the flow and modify it [9]. Usually, it is hopeless to attack the problem in full. It is convenient to describe mixing in terms of passive interfaces and then add, at small scales where the consequences are most important, the effect of active microstructures [10]. This corresponds to the study of a droplet or microstructure in a Lagrangian flow around a material point, i.e.,

\[
\dot{x} = (\nabla \psi) \cdot x + \text{higher order terms}
\]

where \( x \) measures relative distances with respect to the center of mass of the microstructure and \( \psi \) the relative velocity.
A material filament of length $dx$ and orientation $m = dx/|dx|$ evolves according to

\begin{align*}
D(ln\lambda)/Dt &= (\nabla \mathbf{v}) : \mathbf{m}m
\quad (4.1) \\
Dm/Dt &= (\nabla \mathbf{v}) . m - (D : \mathbf{m}) m
\quad (4.2)
\end{align*}

where $\nabla \mathbf{v} = D + \Omega$, $D$ being the stretching tensor, $\Omega$ the vorticity tensor, and $\lambda$ the stretching ratio ($\lambda = |dx|/|dX|$, $dX$ is the initial length of the filament). A similar viewpoint can be adopted in the case of active microstructures. If the element has some internal resistance, it stretches and rotates in a different way than a passive element; an active element obeys a somewhat more complicated expression than (4.1-2). Models can be developed by focusing on the fluid mechanics in the neighborhood of the microstructure [11].

There are almost no experimental data studying the deformation of droplets with a prescribed deformation history (i.e., a specification of $\nabla \mathbf{v}(t)$). Also, a general theoretical analysis is difficult and simplifications are necessary. For example, in [12], we developed a model suitable for low viscosity drops with pointed shapes of viscosity $\mu_i$ immersed in a fluid of viscosity $\mu$ with interfacial tension $\sigma$ and with orientation $m$ and length $l(t)$. The behavior of the droplet in the linear flow field (3) is given by

\begin{align*}
Dlnl(t)/Dt &= (\nabla \mathbf{v}) : \mathbf{m}m - (\sigma/2(5)^{1/2}a)(l(t)/a)^{1/2}(l + 0.8p(l(t)/a)^3))
\quad (5.1) \\
Dm/Dt &= (GD + \Omega).m - (GD : \mathbf{m}) m
\quad (5.2)
\end{align*}

where $G(t) = (1 + 12.5 a^3/l(t)^3)/(1 - 2.5a^3/l(t)^3)$, $a$ is the radius of a spherical drop of the same volume, and $p = \mu_i/\mu$, is the viscosity ratio which is assumed to be very small. The underlined term in (5.1) acts as a resistance to the deformation. Several other models are possible. An equation useful for a variety of axisymmetric elements, such as droplets or macromolecules, is

\begin{align*}
D\mathbf{R}/Dt &= (GD + \Omega).\mathbf{R} - G(F(l+F))(D:rr)\mathbf{R} - [\alpha(l+F)]\mathbf{R}
\end{align*}

where $r = \mathbf{R}/|\mathbf{R}|$, $\mathbf{R}$ is the length of the element, and $G$, $F$, and $\alpha$ are suitably selected parameters [11].

The long range question is what happens to microstructures placed in a chaotic flow. Besides the obvious implications to mixing, such a question has relevance in various contexts; for example it is relevant to non-Newtonian effects due to small concentration of polymers and to drag reduction in turbulent flows. In principle, we can gain some understanding of this problem by means of computer simulations or experiments. Both kinds of studies are probably necessary, even though the connection between the two probably has to be of a qualitative nature. One possibility is comparison of
mixing experiments involving passive and active blobs (with the same initial conditions) and conducted under identical mixing prescriptions (e.g., 2.1-2). Such experiments, if carried out in a chaotic flow field, might magnify differences in the dynamics of the microstructures and produce macroscopic changes in the result of mixing experiments such as those of Figure 1. We anticipate that the differences will be magnified if the timescale of the strain fluctuations in the chaotic regions are of the same order of magnitude as the relaxation time of the microstructures. However, before such a study is undertaken we have to understand as much as possible about the behavior of a single microstructure element in a prescribed kinematics. An indication of the expected results for periodic flows is given by the studies of Nollert and Olbricht [13]. Other results for time dependent flows indicate the severity of transient effects. For example, in [14], we considered the number of fragments $N$ produced by bursting of an infinite thread by means of a model based on growth of capillary waves. The flow field around the thread was taken to be

$$v_z = \Gamma(t) z, \quad v_r = (-1/2) \Gamma(t) r .$$

Two cases were compared; one with constant $\Gamma$, the other with $\Gamma(t) = S/(1 + St)$ where $S$ is a constant. We obtain that $N$ goes as $E^{2.8}$ for constant and as $E^1$ for the case $\Gamma(t) = S/(1 + St)$ (where $E$ is the capillary number, $E = \mu Sd/\sigma$, where $\mu$ is the viscosity of the continuous medium, $d$ the initial diameter of the thread, and $\sigma$ the interfacial tension). This implies that under identical conditions a flow with decaying efficiency will produce significantly fewer fragments than one with constant efficiency. The next studies should focus on breakup in sequences of weak flows (see [14]). More detailed models might incorporate coalescence and redispersion and the interaction and percolation of significantly stretched regions in space.

ACKNOWLEDGMENT

This work was performed under the auspices of the U.S. Department of Energy under grant DE-FG02-85-ER-13333.

REFERENCES


[8] i.e., by following an initial blob composed of many particles and integrating the equations of motion (1) to compute the location of the points avoiding crossing of lines and taking into account the exponential loss of precision of initial conditions.


**Figure 1.** Mixing of a blob of a tracer in a low Reynolds number cavity flow. The fluid is glycerine and the tracer a dilute solution of a fluorescent dye with low diffusion coefficient dissolved in glycerine; the cavity is approximately the same size as the figure. Illumination is provided by ultraviolet light. The system follows the prescription of Equations (2.1-2) with a period $T$ of 30 seconds. The top figure represents the system after $31/2$ cycles whereas the bottom figure represents the system after 7 cycles. Note the almost exponential mixing growth and the "holes" of unmixed material (i.e., the dye does not invade these regions). Note also that the large scale features (folds) are the same in both figures, the structure is apparently set by low period horseshoes. The large (top) hole translates and deforms with repeated applications of the cycles but it does not disappear even for long times (from K. Leong, PhD thesis in progress, Univ. Massachusetts, Amherst).
COMMINUTION WITHIN PARTICLE BEDS

Carl R. Peterson
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT

A four-part, analytical and experimental study of comminution within particle beds is described. Both single particle behavior and bed internal behavior are treated, for spherical particles. A 2-dimensional computer graphics simulation of bed behavior within moving boundaries is presented, including fragmentation of individual particles when specified failure criteria are met. Bed behavior may be described as the repeated build-up of internal, frictional, load-bearing structures, followed by major collapse triggered by the failure of one or a few key particles. The simulation is compared to experimental bed behavior with encouraging results. Particle failure is explored both experimentally and analytically, and it appears likely that an "hydrostatic" stress condition brought on by multiple contacts (yet to be explored) will be of major importance.

INTRODUCTION AND BACKGROUND

Liberation of separate minerals, for applications from the cleaning of coal to beneficiation of ores, requires fine grinding of the raw material. Comminution, the reduction of material to smaller pieces, is thus an increasingly important process as demands for deep cleaning of coal increase and domestic ore grades decline. The process is a major energy consumer: 2% of all U.S. electric power generation goes into comminution (5% world-wide). At the same time, there is room for improvement: some say that current comminution devices are less than 1% efficient [1]. With such incentive and apparent room for improvement, there has been considerable research on comminution, but with little major effect, at least in overall process efficiency. While energy consumption, per se, is momentarily not of great concern, low efficiency has other important consequences. Energy input which does not go into comminution of feed material must go somewhere, mostly to noise, heat, and machine wear, and wear is of major concern.

Much of the more scientific research on comminution has considered the fragmentation of individual particles. At least three "theories" have been advanced to predict the energy required to fracture a particle of characteristic dimension "D". In early work Rittinger proposed that the energy necessary is proportional to $D^2$, in keeping with the energy necessary to create new surfaces, while Kick proposed that it should be proportional to $D^3$, in keeping with a volume energy storage. More recently, Bond [2] expressed his "third theory of comminution", which selects 2.5 as the exponent, largely because it is midway between the others. Peterson [3] has suggested that the $D^2$ relationship is correct for very small particles while $D^3$ is correct for larger particles, based on surface-to-volume ratio relationships and physical reasoning that was also separately noted by Schonert [4]. Others have suggested a range of exponents, including variable exponents to fit empirical curves [5,6]. Larson [7] presents a recent summary of these relationships.
Though an understanding of single particle behavior is critical, practical comminution devices, at least of conventional and reasonably foreseeable types, never handle single small particles: they process very large numbers of particles in particle beds. This study integrates experimental and analytical approaches, and both single particle and particle bed behavior. The intent is to develop an understanding of particle fracture within particle beds: what particle properties determine its fracture loading, and what bed conditions best create that loading? With such an understanding of behavior within beds, even if only qualitative, it is believed that more efficient comminution devices can be designed. Key to development of this understanding is development of a two-dimensional, computer graphic simulation of bed behavior, including a determination of the loading on and motion of individual particles, and simulation of particle failure when loading reaches appropriate failure criteria. A correct picture of internal behavior would provide valuable design guidance, both in stimulating new concepts which create favorable bed conditions and in checking proposed new designs.

Early in the present program it became clear that it would be necessary to simplify the study by restricting it to spherical particles. Granular particles, while obviously of greater practical significance, introduce an infinity of variables in shape and loading/failure conditions. Even "simple" experimental work, like loading a particle to failure between flat surfaces, is complicated by multiple orientation possibilities and by the lack of distinct failure criteria. For example, does local failure of a sharp corner constitute particle failure? It is believed that a bed of randomly sized spherical particles will provide the sought-after understanding and guidance. Once a clear picture of spherical particle bed behavior emerges, however, this assumed qualitative similarly must be confirmed. With this introduction, the four areas within the program are discussed separately, although it will be evident that all are closely interrelated and essential to a complete understanding. This paper summarizes thesis work recently completed at M.I.T. by Arnold Larson, Kathleen Laffee, Gary Drlik, and Zhuo-Ru Ding, references [7] through [10] respectively, and continuing thesis work underway by John Pflueger and Adriana Praddaude. The work has been funded by the Department of Energy and by an industry consortium, the Synfuels Program of M.I.T.'s Energy Laboratory.

SINGLE PARTICLE WORK, EXPERIMENTAL

This work set out to explore the failure of single particles and to provide guidance on the selection of and quantitative behavior of materials to be tested in particle beds [7]. It clearly indicated that irregular granular shapes were, in themselves, hopelessly complex and unsuited to bed behavior studies (at least initial studies). A small compression test machine was constructed, with automated data acquisition providing force-deflection curves, peak force, and energy to fracture. Materials tested included glass, alumina, ammonium nitrate, and calcium chloride. Most tests were two-point contact between flat tungsten carbide anvils. In response to a growing realization that multiple-contact loading may be of major importance, as discussed in following sections, some (largely inconclusive) three-and four-point loadings were tried between "V"-grooved anvils. Typical data for two-point loadings are shown in Figure 1. Considerable scatter is evident, even for this simplest shape and "uniform" material. From such tests on a variety of materials it was decided that glass spheres would be best for early bed testing because failure, although occurring at varying loads and energies, is sharp and distinct.

This apparatus will see continuing use to quantify individual particle behavior in the search for appropriate failure criteria. As might be expected, adequate representation of failure loading is essential to realistic bed simulation.
PARTICLE BED BEHAVIOR, EXPERIMENTAL

This work provides experimental data against which to compare simulation behavior [8]. Although a range of geometries will be used, work to date has been limited to simple compression between parallel flat plates, with the top plate moving downward toward a stationary lower plate. A variety of spherical materials was tested, but as noted, glass spheres were selected for most of the work. Nominal 2.0 mm glass spheres were used, with actual sizes varying between 1.5 and 2.5 mm. Loads of up to 20,000 pounds were applied on circular sample beds of 1.5 inches diameter, and of varying bed depth (usually 1.5 inches). Force-deflection curves were obtained and bed material was screened before and after tests to determine particle size distributions. A novel casting technique was developed to "freeze" the damaged bed in place and, by sectioning the cast bed, determine where damage actually occurred within a bed.

It would not be difficult to develop a simulation which simply indicated increasing load with increasing bed compression. To develop confidence in the simulation of bed behavior, sufficient for use as a design guide, more detailed or "second order" behaviors were sought. Two such behaviors were demonstrated. The first, shown in Figure 2a, was a load-deflection curve which demonstrated an unexpected, but highly repeatable, drop in load with continuing compression. The repeatability of this phenomenon is not well understood, but it may be due to a sudden shift from irregular to more regular packing when early structures within the bed fail due to the failure of key particles. Similar behavior is observed in simulations as discussed in the following section, but it is not clear that the simulations are as repeatable. The second, also repeatable, behavior is a tendency for particle failure to begin at the top of the bed and to progress downward as compression continues. This can be seen in Figure 3, and it was confirmed by separate screening of different bed levels. Some hint of similar top-down failure can be seen in simulations, but it is not yet clear that this is more than an edge effect.

Both behaviors are believed to be due to variations in packing geometry throughout the bed. Constrained by the flat bottom surface, packing near the bottom is quite regular, but it becomes increasingly random with increasing depth because the spheres are not of uniform size. Apparently the random packing suffers irregular load distribution with higher local loads (even though the top of the bed was flat before load application), leading to early fracture and a weaker overall structure. To test this hypothesis, a bed was carefully prepared to achieve greater order throughout, and a stronger structure resulted, as seen in Figure 2b. Conversely, a weaker structure is created when a rough lower surface (created by a screen insert) breaks up the regular
packing there, as seen in Figure 2c. In other tests, the observed load drop is
reduced as an increasing fraction of smaller (0.3mm) spheres is mixed in, preventing
shifting of the structure of 2mm spheres (or the scale of the drop may simply be
reduced so that it cannot be separated from the irregular higher frequency load varia-
tions).

Figure 2 Force-deflection curves for compression of beds. [8]

Figure 4 indicates the extent of particle damage as a function of energy input
(integration under the force-deflection curve) for two different screen sizes. The
percent not retained is simply the percent of spheres broken from the initial (zero
energy) input sample. Varying energy inputs were achieved by stopping compression
at the desired point and screening the resultant bed. The linear energy-damage relation-
ship suggests that there are no significant changes in efficiency over the various
portions of the repeatable load-deflection behavior.

TWO-DIMENSIONAL BED SIMULATIONS

This work provides a two-dimensional picture of the internal behavior of particle
beds of circular "particles". Apart from the obvious simplification, a two-
dimensional solution is actually desired at this stage to provide a clear visual
understanding. The first work [9] utilized a purchased program for assemblies of cir-
cular particles within moving boundaries. The program essentially solves transla-
tional and rotational equations of motion for each particle subject to elastic normal
forces and frictional forces at particle-to-particle and particle-to-boundary con-
tacts. To this was added a failure criterion in terms of the calculated particle con-
tact forces and a model for "broken" particles. A second program, under development
by John Pflueger, is more convenient and allows greater flexibility. In both cases,
relatively small beds of 100 to 150 initial particles are treated (compared to several
thousand in experimental work), although the number increases to 500 or more as "frag-
ments" are created.

A simple failure criteria is sought and two (perhaps too simple) have been used to
date. Average stresses along a diameter, defined as that stress necessary to equili-
brate the external forces on one side, are determined for all diametral orientations,
and "average principal stresses" found. The first failure criteria simply assumed
failure when the maximum principal stress reached a set value. The second criteria, probably more realistic, indicates failure when a set shear stress, or difference between principal stresses, is reached.

Figure 3  Sections of crushed bed, 2 mm glass spheres. [8]
The model for "broken" particles, shown in Figure 5, is roughly based upon creation of major fragments under two contact forces. The model is convenient since all four "fragments" will always fit within the space vacated by the original circle. "Material" lost in this model is shaded in Figure 5. At least initially, this loss might be ascribed to the creation of fine material that fits within the interstices between larger particles, but for greater bed compression the important constraining effect of this material cannot be ignored. Assuming that fragmentation is initiated at a single contact, the cluster of four new particles is aligned in the bed with the largest contact force. Fragments are assumed circular since, with the initial assumption of a bed of spherical particles, there is no other meaningful particle shape in this simulation.

Simulations indicate that both the form of the particle failure criteria and the assumed friction coefficient are important variables. Figure 6 illustrates four cases that all started with the same bed (a nicety not available experimentally) and all compressed to the same degree. Comparing low friction ($\mu = .05$) and high friction ($\mu = 1.0$) cases, it can be seen that higher friction results in greater damage. Comparing the two failure criteria, it appears that the maximum shear stress criterion favors early failure at the bed boundaries (a difference that is more evident in other simulations such as seen in Figure 8). It is hoped that comparisons such as these will evolve "correct" parameters with which the simulation duplicates experimental observations in detail, in which case the simulation can be used as a powerful design tool in search of more effective geometries.
Figure 6  Particle damage comparisons. [9]

Figure 7, from the more recent simulation, illustrates the vector contact forces within the bed, with line thickness proportional to force magnitude (forces less than 10% of the maximum are not shown). Note the non-normal forces, indicating appreciable frictional components ($\mu = .7$). Distinct, load-bearing structures are clearly evident in the left figure. Failure of a single particle, at "A", leads to the virtual collapse of the entire structure and a significant load drop as the top boundary moves down at constant velocity*. This build-up and collapse of frictional structures is repeated many times as the bed is compressed.

Figure 7  Load-bearing structures within bed. [Pflueger]

* Particles have also failed in the bottom corners, but their influence on overall load is probably not significant.
Figure 8 presents load-deflection and total bed population curves. Load variations accompanying transient internal structures can be seen. With the top wall descending at a constant rate, "calculation cycles" are the same as time or wall displacement. Since failure of a single particle creates four new ones, each failure results in an increase of three in the total number of particles (discs) in the bed. Both simulations exhibit an initial rise in force followed by a substantial drop, reminiscent of the experimental behavior shown in Figure 2, but two different explanations are possible. In 8a, a low friction case ($\mu = 0.05$), the load drop is accompanied by progressive fracture of a large number of particles, whereas in 8b, a high friction case ($\mu = 1.0$), the load drop seems to be caused by a shifting of particles triggered by fracture of relatively few. The validity of the simulated behavior, its reproducibility, and the influence of friction coefficient have not yet been explored. While these curves happen to represent the maximum normal stress failure criterion, similar simulated behavior is observed for the maximum shear stress criterion.

![Figure 8 Simulated force-deflection curves.](9)

SINGLE PARTICLE BEHAVIOR, ANALYTICAL

Simulations indicate that, even without consideration of the internal details of fracture propagation, the proper form of the particle failure criterion is of critical importance. Accordingly, a preliminary analytical study of particle stresses and fracture tendency subject to multiple loading has been completed [10]. This study includes both the geometry and load locations for certain idealized bed packing arrangements and the approximation of internal stresses under multiple loading.

In view of other complications, a simple failure criterion is highly desirable, but the present models may be too simple. It has long been understood that the presence of excessive fine material within the bed tends to distribute loads between many contacts, effectively blunting the application of external load on the bed. In addition, this study suggest that excessive contacts can approach an hydrostatic stress within a particle, making fracture impossible no matter how high the individual loads. This finding, if true, can have practical significance: the "blunting" effect might be overcome with additional external load, while the "hydrostatic" effect would not. For this reason the maximum shear stress criterion is believed to be the better of the two criteria used.

The behavior of particle "A" in Figure 7 illustrates the consequences of this model. Particle "A", subjected to just two major forces, fails while adjacent particles, each subjected to three evenly spaced similar loads, do not. Figure 9 quanti-
fies this differing behavior. Disc C, subjected to loads which create high shear stress, will fail, while similarly sized Discs A and B, subject to higher individual loads but lower average shear stresses, do not. Obviously this example does not prove the failure hypothesis - it merely illustrates its consequences.

This behavior was also demonstrated experimentally by mixing 100 2 mm particles in a bed of 0.3 mm particles. The load-deflection curve, seen in Figure 2d, shows little or no load drop. There was no damage to any of the larger spheres, even at loads to 11,000 pounds, whereas 9,000 pounds created considerable damage in a bed of only 2 mm spheres.

For a more thorough examination of this important behavior, it is hypothesized that fractures are initiated at a single contact, but that fracture propagation is strongly influenced by the presence of other loads. Figure 10 is a photo of a fracture that was initiated in a 2 mm glass sphere, but which did not propagate. There were many indications of such damage without fragmentation.

A continuation of this work, including finite element analyses of fracture propagation with and without lateral loads, and experimental study of the same phenomena, is key to the development of a realistic bed simulation. It may be that excessive fine material is far more detrimental to efficient comminution than the recognized "blunting" effect may suggest.
CONCLUSIONS AND RECOMMENDATIONS

The four-part experimental and analytical examination of comminution within particle beds represents a minimum set of essential parallel approaches. Findings to date are encouraging and self-consistent, but far from conclusive. A picture is emerging of frictional structures within the bed, building and collapsing, causing some particles to fail while protecting others. Behavior is dependent upon particle failure mechanisms, frictional properties, and bed packing geometry as well as other variables. With sufficient confidence in the validity of this picture, the resultant understanding can be used as a design guide to maximize those bed conditions which lead to failure. It is recommended that research continue in all four areas.

BIBLIOGRAPHY

APPLICATION OF ACOUSTIC EMISSION TECHNIQUES TO CRYOGENIC EXPERIMENTS

Y. Iwasa
Francis Bitter National Magnet Laboratory and Plasma Fusion Center
and
Department of Mechanical Engineering
Massachusetts Institute of Technology, Cambridge, MA 02139

ABSTRACT

This paper presents the application of acoustic emission techniques in cryogenic experiments. The techniques have proven effective in diagnosing superconducting magnets: they can also be used to quantify dissipative energies released in tensile and fracture experiments.

INTRODUCTION

The application of acoustic emission (AE) techniques for monitoring superconducting magnets includes: (1) identification of quench-causing mechanical disturbances; (2) quantitative relationship between motion-induced acoustic activity and an analytical model based on frictional motion in the magnet winding; (3) normal-zone detection; and (4) energy quantification.

Acoustic emission monitoring of superconducting magnets was first done in the late 1970s. It has now been firmly established that acoustic signals in superconducting magnets are emitted principally by mechanical events such as conductor strain, conductor motion, frictional motion, and epoxy cracking. These mechanical disturbances are transitory, each generating a packet of signals that can be triangulated with sensors strategically located.

ACOUSTIC EMISSION IN SUPERCONDUCTING MAGNETS

Quench Identification

By combining the AE and conventional voltage measurement techniques, it has become possible to identify and localize the source of an event responsible for a premature quench. Generally, quench events in epoxy-impregnated superconducting magnets may be divided into three classes.

Class 1: a quench preceded by a voltage spike and an AE signal. Because sudden conductor motion generates an AE signal and induces a voltage spike across the coil, the simultaneous occurrence of a voltage spike and an AE signal implies that such a quench to be conductor-motion induced. Figure 1 presents an oscillogram showing an example of a conductor-motion-induced quench event observed in a long superconducting dipole, 1.8 m long. The top two traces are AE signals and the bottom trace is the coil terminal voltage. Sensor AE 2 was located at one end and sensor AE 4 at the other end of the dipole. The voltage pulse that precedes a slowly rising resistive voltage was induced by the conductor-motion event, most probably a microscopic conductor motion. The event was first
recorded by sensor AE 4 and then, after a delay of 0.3 ms, by sensor AE 2. Since the propagation speed for the acoustic wave is \( \sim 5000 \text{ m/s} \), the delay time of \( \sim 0.3 \text{ ms} \) suggests a wave-travel distance of \( \sim 1.5 \text{ m} \), or that the event did indeed take place near one end of the dipole. This is a good example of an acoustic-emission technique used not only for quench source identification but also quench source triangulation in a superconducting magnet.

Another example of class-1 quench, recorded in an Nb\(_3\)Sn solenoid operated in a bath of superfluid helium at 1.8 K, is shown in Fig. 2. Note both the acoustic signal and the voltage spike at the start of the quench, which occurred at 520 A.

**Class 2:** a quench with an AE signal but no voltage spike. This quench is triggered by a non-conductor-motion event such as epoxy cracking. Figure 3 corresponds to an epoxy cracking event triggering a quench in an epoxy-impregnated coil. The left oscillogram (Fig. 3a) is similar to the oscillogram of Fig. 1 except here there is no motion-induced voltage spike (bottom trace). Note that a resistive voltage grows \( \sim 16 \text{ ms} \) after the cracking event. The two acoustic emission traces shown in an expanded time scale \( (\sim 25 \mu\text{s per division}) \) in Fig. 3b clearly indicate that the event occurred at near the top flange of the coil. The AE sensor for the top trace was mounted on the top flange and the sensor for the bottom trace was mounted on the bottom flange. The distance between the flanges was 109 mm. A time delay of \( \sim 50 \mu\text{s} \) (two divisions) gives rise to an expected propagation speed of \( \sim 2.000 \text{ m/s} \) for the AE wave through the coil winding.
Class 3: a quench with neither a voltage spike nor an AE signal. Class 3 quench results from joule heating as the conductor reaches a critical current and enters the normal state, the desired goal for any superconducting magnet. Because the quench is induced by pure heating rather than by a mechanical event, no voltage spike or AE signal appears at the start of the quench. Figure 4 presents an example of class-3 quench (at 540 A) recorded in the same Nb$_3$Sn solenoid discussed in connection with Fig. 2 result. Note the absence of both AE signal and voltage spike.

This AE/voltage technique has been successfully applied to identify, and sometimes localize, quench sources in many magnets.$^{10-14}$

Conductor-Motion-Induced AE Signals in a Superconducting Dipole

Acoustic emission results obtained in a superconducting dipole at the Fermi National Accelerator Laboratory are presented below.$^{12}$ The data have shown that most AE signals in these dipoles are induced by conductor motion in the winding. The motion-induced acoustic emission data are correlated with a simple friction model. Agreement between data and theory is excellent.

To monitor specifically superconducting dipoles and quadrupoles of high-energy accelerators, we have developed an instrumentation system capable of (1) real-time location of genuine AE events from several sensors and (2) real-time extraction of AE energy. The system was field tested to monitor a model dipole at Fermi National Accelerator Laboratory (FNAL).

Experimental Procedure: The sensors were the differential type. Each sensor contained two piezoelectric elements cut from one circular disk, one element oriented positively with respect to ground and the other negatively. The differential sensor has higher sensitivity and signal-to-noise ratio than the single-ended type. The two halves were enclosed in a copper jacket which shielded them from electromagnetic interference. The sensor was then mounted on the magnet collars with a G-10 piece fixed on either end by studs welded to the collars and nuts. The sensors were further protected from physical abuse by diecast aluminium guard boxes. Belleville washers acted as stiff springs preventing the guard boxes from loosening on cooldown.

Amplification was accomplished in two stages: first, with a commercial preamplifier with a good common mode rejection and wide bandwidth (2–600 kHz); the preamplifier gain was selectable at 40 or 60 dB. Further amplification was done in the secondary amplifier (part of the signal processing system). At FNAL, the preamplifiers were located immediately outside the magnet test-rig and connected to the sensors by 3-m coaxial cables. The preamplified signals were carried over 30 m of shielded cable to the signal processing system located in the control room. The signals usually picked up noise (mainly low-frequency power supply noise or its higher harmonics) along the path to the control room. Active high-pass filters with cut-off frequency of 10 kHz were inserted between the preamplifier outputs and the signal processor.

Four sensors are used in all, one each at the top, middle and bottom of the magnet. The fourth sensor is a guard, mounted on the current leads to detect extraneous noise. The signal processor uses the first hit method to detect events. Event count, energy count, and time of arrival are digitized and transferred to the computer for storage and further processing.
Results: Only those results relevant to the friction model are presented here. Quench identification and localization results are not included. From the principal set of data (cumulative record of acoustic emission events, energy and transport current history), only cumulative energy count versus transport current I data are presented and compared with the prediction of the friction model theory. Given a motion-induced acoustic emission signal, $v_{AE}(t)$, acoustic energy $E_{AE}$ is defined as:

$$ E_{AE} = \int v_{AE}^2 dt. \quad (1) $$

For each motion-induced AE signal arriving first at one of the three sensors, $E_{AE}$ was computed and cumulated for each sensor as the dipole current was raised from zero to 5000 A.

Of several results obtained, it was recorded that there was substantially more acoustic emission activity (and thus greater $E_{AE}$) in the virgin run than in any subsequent runs. The virgin run is that when the dipole is energized for the first time: the virgin run returns when the dipole is driven normal, heated substantially and subsequently returns superconducting. Later runs, on the other hand, did not differ appreciably from each other. Figure 5 presents cumulative $E_{AE}$ versus $I$ plots recorded by the bottom sensor for the virgin run (trace 1) and nonvirgin runs (traces 2, 3, 4). The dotted lines are theoretical based on the frictional conduction-motion model.\textsuperscript{12}

Frictional Conductor Motion Model

Analytical expressions for $E_{AE}$ versus $I$ may be developed, based on a simple frictional mechanism controlling conductor motion within the dipole winding.\textsuperscript{15} Assume that the conductor is subjected to three main forces: spring, electromagnetic ($f_s$), and friction ($f_c$). The force balance for a unit conductor, displaced by distance $x$ from the stationary position, is then given by:

$$ kx = f_s - f_c. \quad (2) $$

where $k$ is the spring constant. Static friction $f_c$ may be expressed as:

$$ f_c = f_s \quad \text{if } f_s > kx \quad (3a) $$

$$ f_c = -f_s \quad \text{if } f_s < kx \quad (3b) $$

The term $f_s$ is a positive value. The sign change occurs because friction tends to oppose motion.

Fig. 5 Cumulative $E_{AE}$ vs. $I$ plots recorded for the virgin run (trace 1) and nonvirgin runs (traces 2, 3, 4). The dotted lines are theoretical based on the frictional conduction-motion model.\textsuperscript{12}

Fig. 6 Force-displacement contour traced by conductor.\textsuperscript{12}
The conductor begins in the virgin state and zero transport current (point A in Fig. 6). As the transport current $I$ is increased, $f_e$ increases. Initially, the whole conductor remains still, emitting no acoustic signals. This corresponds to the vertical trajectory beginning at A. When $f_e$ reaches a threshold value $f_o$ small sections of the conductor where $f_e$ is least begin to move. This is point B on the $f - x$ space.

Beyond B, the energy released ($\delta E$) when the conductor moves a distance $\delta x$ due to excess electromagnetic force over friction $f_e - f_o$ is given by:

$$\delta E = (f_e - f_o) \delta x. \quad (4)$$

Substituting $\delta f = kf \delta x$ and $\delta f = \delta f_e$ (since $f_o$ is constant) and integrating, we obtain an expression for the total energy $E$ released as $f_e$ is increased:

$$E = \int_{f_o}^{f_e} \frac{(f_e - f_o)}{k} df_e = \frac{(f_e - f_o)^2}{2k}. \quad (5)$$

Assuming a linear system, the energy picked up at the sensor will be:

$$E = \frac{3(f_e - f_o)^2}{2k} \propto E_{AE}. \quad (6)$$

The $3$ term is a proportionality constant which depends on the transfer properties of the medium, the sensitivity of the sensors, and the calibration factor of the energy processor.

Equation 6 holds until point D when $I$ begins to decrease. So far the conductor has moved outwards. When $I$ begins to decrease the conductor will tend to move inwards; however, the frictional force reverses direction and the conductor remains still until point G when it begins to move. The value of $f_e$ at point G is $f_m - 2f_o$, where $f_m$ is the value of $f_e$ at D. Using similar reasoning to that above, the energy released during sweep-down as $f_e$ decreases from $f_m - 2f_o$ is:

$$E = 3 \int_{f_o}^{f_e} \frac{|f_e - (f_m - 2f_o)|}{k} df_e = \frac{3|f_e - (f_m - 2f_o)|^2}{2k}. \quad (7)$$

This equation holds until the current reaches zero at point F where $f_e$ becomes zero and the conductor stops moving. The displacement at this point is $x_o = f_o/k$.

During nonvirgin runs the trajectory starts from F. The conductor remains still and no AE is emitted until C when $f_e = 2f_o$. Thereafter, the conductor moves, thus generating acoustic signals. Following the same steps as before we find that the total energy released is given by:

$$E = \frac{\beta(f_e - 2f_o)^2}{2k}. \quad (8)$$

In the return passage, the trajectory follows path DGF as in the virgin run. During nonvirgin runs the trajectory always follows path FCDGF with reproducible acoustic characteristics.

**Theoretical Predictions:** The electromagnetic force on a noninductive winding with a transport current $I$ is proportional to $I^2$:

$$f_e = \alpha I^2, \quad (9)$$

where $\alpha$ is a constant. Substituting appropriately into Eqs. 6, 7, and 8, the following expressions are
obtained for the predicted acoustic emission energy as a function of $I$ for the virgin run (Eq. 10) and the nonvirgin run (Eq. 11):

$$E = \begin{cases} 
0 & \text{if } 0 < I < I_o \\
\eta(I^2 - I_o^2)^2 & \text{if } I_o < I < I_m \\
\eta(I_m^2 - I_o^2)^2 & \text{if } I_m > I > \sqrt{I_m^2 - 2I_o^2} \\
\eta(I^2 - (I_m^2 - 2I_o^2))^2 & \text{if } \sqrt{I_m^2 - 2I_o^2} > I > 0
\end{cases}$$

Sweep

Up

Sweep

Down

(10)

$E = \begin{cases} 
0 & \text{if } 0 < I < \sqrt{2}I_o \\
\eta(I^2 - 2I_o^2)^2 & \text{if } \sqrt{2}I_o < I < I_m \\
\eta(I_m^2 - 2I_o^2)^2 & \text{if } I_m > I > \sqrt{I_m^2 - 2I_o^2} \\
\eta(I^2 - (I_m^2 - 2I_o^2))^2 & \text{if } \sqrt{I_m^2 - 2I_o^2} > I > 0
\end{cases}$

Sweep

Up

Sweep

Down

(11)

$I_o$ is the current at which the coil begins to move in the virgin run and it is given by $\sqrt{I_o/\alpha}$ and the constant $\eta$ is given by $(\beta \alpha^2)/(2k)$. Note that Eqs. 10 and 11 are the same as those obtained by the previous analysis\textsuperscript{15} where it was assumed (1) that the conductor moved in discrete steps, each step corresponding to one acoustic emission event, and (2) that the number of oscillatory signals above a given threshold was proportional to the distance travelled during the step motion. In deriving Eqs. 10 and 11, however, the only assumption made was that the system was linear. The theoretical predictions of Eqs. 11 and 12 (using the values of $I_m$ and $I_o$ obtained from the experiments) are plotted by the dotted lines in Fig. 5.

Figure 7 is an example of data taken while the magnet was ramped continually up and down, well below its quench current. As predicted by theory, starting at zero current there are initially no acoustic signals until a certain current level is reached, where acoustic emission begins to occur, the rate rising to a peak coincidentally with the transport current. When the current stops and starts decreasing, the static friction reverses direction, temporarily stopping the coil from moving. The AE rate will remain zero for a while after the current starts decreasing, until static friction is overcome. The AE rate must also be zero at the end of the sweep-down when $I = 0$. In between, the $E_{AE}$ rate goes through nonzero values and must thus have at least one peak. This second peak, which occurs at current $I_p$, may be predicted from Eq. 11. Namely:

$$E = \eta (I^2 - (I_m^2 - 2I_o^2))^2$$

(12)

and

$$\left( \frac{d^2E}{dT^2} \right)_{I_p} = 0$$

(13)

The corresponding current is then given by:

$$I_p = \sqrt{\frac{(I_m^2 - 2I_o^2)}{3}}$$

(14)

For the case shown in Fig. 7, $I_p$ predicted by Eq. (14) is 1400 A, which agrees reasonably well with the observed value of 1600 A.
ENERGY QUANTIFICATION

Since each of mechanical disturbance events such as those taking place in superconducting magnets is shown to emit an AE signal, $v_{AE}$, it is possible to use an energy parameter $E_{AE}$, defined in Eq. 1, to quantify event energy. The validity of Eq. 1 was more directly tested in a series of tensile tests performed on copper and other materials. It was observed that copper (annealed OFHC) at 4.2 K in the plastic regime elongated essentially as a series of steps, each 0.5-1.5 $\mu$m. A 1-$\mu$m elongation corresponds to a strain of $\sim 1 \times 10^{-3}$. This stepwise elongation, absent when the copper specimens are tested at room temperature and 77 K, is believed to be associated with twinning. Figure 8 presents dissipative energy data in which each $F_t \Delta x$ represents the mechanical work supplied to the specimen by the experimental rig and dissipated as heat during one "microslip" event; $E_p$ is the corresponding dissipation energy measured directly by an energy transducer, specially designed for the experiment. Note that there is an excellent one-to-one (45°) correspondence between $F_t \Delta x$ and $E_p$. Presented in the same figure are $E_{AE}$ data, each computed from $v_{AE}$ recorded for each microslip event. Although $E_{AE}$ cannot provide energy information on an absolute basis, the 45° line drawn through $E_{AE}$ data suggests $E_{AE}$ to be a useful parameter in quantifying dissipation energies, particularly in the range ($< 100$ $\mu$J) where the direct measurement is very difficult.

CONCLUDING REMARKS

Since their introduction ten years ago for monitoring superconducting magnets, the acoustic emission techniques have proven very effective in identifying the source of premature quenches in high-performance superconducting magnets: conductor motion and cracking of epoxy materials used to fill the coil winding have been identified as the two principal sources. The techniques have been also applied to quantify microscopic dissipative energies released in tensile and fracture experiments. At the present time the most critical quality needed in AE technology before it can receive a wider usage in cryogenic experiments is reliability. To achieve this ultimate goal of wide usage, it is necessary to advance understanding in the following areas: 1) sensor sensitivity, particularly in cryogenic temperatures; 2) AE wave propagation in the magnet winding media; 3) quantitative correlation between AE amplitude and physical quantities such as displacement, velocity, force, and energy. This last area has proven most difficult and been least successful in acoustic emission technology in general. It is also important to realize that acoustic emission techniques are best served as complementary rather than exclusive monitoring techniques.

ACKNOWLEDGEMENT

The work reported here was performed under the auspices of the U.S. Department of Energy and the National Science Foundation. The author wishes to express his thanks to many of his colleagues, past and present, particularly E.S. Bobrov, H. Fujita, H. Maeda, O. Ige, and O. Tsukamoto.
REFERENCES


MODELING AND ANALYSIS OF SURFACE CRACKS

David M. Parks
Massachusetts Institute of Technology
Department of Mechanical Engineering
Cambridge, Massachusetts 02139

ABSTRACT

Numerical solutions examining the range of applicability of dominant singularity approaches to ductile fracture of structures are explored, as well as the development of simplified, but accurate analytical models.

INTRODUCTION

The primary success of fracture mechanics has been in simplifying the determination of crack front stress and deformation fields. Actually, the problem is not so much "solved" as it is "avoided" by appealing to notions of asymptotic similarity of crack tip fields as depending on, e.g., the stress intensity factors $K_I$, $K_{II}$, and $K_{III}$ of linear elastic fracture mechanics (LEFM), or the J integral governing the strength of local fields in nonlinear (elastic) fracture mechanics (NLEFM) for quasi-stationary cracks. Providing such single parameter asymptotic characterizations actually dominate the complete near-crack-front fields (including the zone of operative microfracture processes) in two different bodies, then the respective loadings can be scaled to give similar crack fields. For a recent review of NLEFM, see Hutchinson [1]. However, there remain certain issues of critical, fundamental, and engineering significance which are not yet well-addressed. We are concerned with inherent parametric limits of applicability of single parameter (e.g., J-based) approaches to describing crack front fields in ductile engineering structures. Within the parametric "window of applicability" of such approaches, we are also concerned with the development of versatile, cost-effective computational tools which can accurately evaluate the singularity amplitude.

The following sections highlight activities in each problem area. In the next section, results of highly detailed numerical solutions of a particular three-dimensional surface cracked plate loaded in far field tension are presented. Next, further development of the simple line-spring model for approximate analysis of part-through surface-cracked plates and shells is presented.

J2 DOMINANCE IN 3D

In the crack analysis of monotonically loaded bodies undergoing nonlinear (plastic) deformation, it is convenient to make use of an "equivalent" hyperelastic material model coinciding with the elastic-plastic material under proportional stressing. A fairly general phenomenological power law model of nonlinear response in uniaxial tension is: $\epsilon = \alpha \epsilon_0 + (\sigma/\sigma_0)^n$.

Here $\epsilon$ and $\sigma$ are strain and stress, respectively, and material parameters are $\sigma_0$, an effective yield stress; $\epsilon_0 = \sigma_0/E$, (E is Young's modulus) a reference yield strain; $n$, the strain hardening exponent; and $\alpha$, a dimensionless factor.

When such response is tensorially generalized using $J_2$ deformation theory plasticity, small geometry change asymptotic analysis leads, as local cylindrical
distance \( r \) from the crack front approaches zero, to crack tip fields

\[
\sigma_{ij}(r, \theta) \to \sigma_0 \frac{1}{1+n} \frac{1}{r} (J/\alpha_0 \sigma_0 I_n r) \quad \text{and} \quad \epsilon_{ij}(r, \theta) \to \epsilon_0 \sigma_0 \frac{1}{1+n} \frac{1}{r} (J/\alpha_0 \sigma_0 I_n r)
\]

(1a)

where \( \sigma_{ij} \) and \( \epsilon_{ij} \) are dimensionless functions of their arguments normalized by the function \( I_n(n) \). These fields, for plane strain, were given by Hutchinson [2] and Rice and Rosengren [3], and are collectively referred to as the HRR singular fields.

The amplitude of these fields is given by the value of the parameter \( J \). Numerical methods for evaluating \( J \) are available [4]. When the asymptotic fields in eqns. (1a,b) "dominate" the complete fields over distances large in comparison to the crack tip blunting and fracture process zones, it is a natural extension of LEFM to correlate crack extension with \( J \). The asymptotic fields (1a,b) cannot apply too close to the crack tip, since effects of blunting and geometry change have been neglected. An effective crack tip opening displacement (CTOD), \( \delta_c \), can be defined as separation where 45° lines emanating from the crack tip intercept the crack face, and the result is \( \delta_c = d_n(\alpha_0, n) \sigma_0 \), where \( d_n \) is only weakly dependent on \( \alpha_0 \), but varies from about 0.8 for large \( n \) to about 0.3 for \( n = 3 \). The length scale \( \delta_c \) is a measure of the finitely deforming region not accounted for in (1a,b). At distances \( R \) greater than \(-5\delta_c \) from the blunted crack, finite strain finite element solutions [5] show the fields (a,b) do apply. Since (a,b) are asymptotic results, they apply at radius \( R \) in a finite body of crack length \( a \), remaining ligament \( c \), etc. only if \( R/a, R/c \) are sufficiently small. In large scale yielding, the requisite smallness of \( R/a, R/c \) depends on amount of strain hardening, and the crack front constraint of the fully plastic flow field. This is particularly evident at high \( n \)-values and plane strain ligament deformation. In the non-hardening case (formally, \( n \to \infty \)), McClintock [6], demonstrated a broad range of fully plastic crack tip stress and strain-increment states. For single edge notched (SEN) geometries, low constraint is associated with predominant tensile loading of the ligament, while predominant bending leads to high crack tip triaxiality similar to the high-\( n \) limit of eq. (1a).

Several studies [7-9] of fully plastic plane strain deformation have lead to the conclusion that dominance can be expected providing the ratio of ligament \( c \), to crack tip similarity length, \( J/\sigma_0 \) satisfies \( \mu = c/(J/\sigma_0) \geq \mu_{cr} \), where \( \mu_{cr} \) is a "critical" lower limit which varies smoothly with the ratio of ligament tension to bending from \( \mu = 25 \) for pure bending to \( \mu = 200 \) for tension [9]. Understanding of corresponding necessary conditions for three-dimensional crack fronts remains slight. While some detailed mesh nonlinear three-dimensional calculations have been performed with an eye toward assessing dominance [10], the mesh fineness used is far less than in the two-dimensional studies cited [7-9].

Recent work by Delatte [11] addressed the question of HRR dominance in a particular three-dimensional geometry, namely a semi-circular surface crack of relative depth to thickness \( a/t = 0.5 \) in a large plate subject to far-field uniaxial tension. The material was modeled using Ramberg-Osgood deformation theory with Poisson ratio \( \nu = 0.3 \), Young's modulus \( E = 32.5 \times 10^6 \) psi (225 GPa), hardening exponent \( n = 7 \), nominal yield strain of \( \epsilon_0 = 0.002 \), and \( \alpha = 3.1 \). These values are broadly consistent with the behavior of A710 steel as reported by Reuter and co-workers [12].

Fig. 1 shows a coarse finite element mesh (3658 nodes, 632 elements, 10974 d.o.f.) of one-quarter of the overall geometry, where advantage has been taken of two symmetry planes. Within this mesh exists a toroidal region of 8 segments of crack
front, each segment spanning 11.25° of arc, and each segment contains 8 focussed 20-node reduced integration (2 x 2 x 2) isoparametric finite elements wrapped about 180°. At toroidal radius, \( p/a = 0.2 \), the displacements of the coarse mesh were used to drive a radially-refined mesh (2673 nodes, 512 elements, 8019 d.o.f.) of the same crack front segmentation. The solution was obtained using the ABAQUS [13] finite element program on an Alliant FX/8 computer. Of the myriad of data generated by such analyses, we have chosen only a small fraction to convey the sense in which HRR dominance at first applies, then is lost.

Figure 1. Semi-circular crack in a flat plate: geometry and coarse mesh.

Figure 2 shows the evolution of crack front J value as a function of remote applied load level, \( \sigma/\sigma_0 \), for several angular locations, \( \phi \), along the crack front. Here \( \phi = 0 \) is the free surface while \( \phi = 90° \) denotes the deepest penetration at specimen center line. As can be seen, except for the near free surface region (which cannot be accurately modeled with the coarse crack front segmentation), J-values are rather uniform along the crack front. This uniformity is consistent with the uniformity of crack tip blunting measurements made on this specimen [12]. Elastically, the maximum J occurred near the free surface, and as load magnitude increased to fully plastic conditions, a small relative maximum occurred near \( \phi = 30° \), a trend also observed experimentally [12].

The relevance of these J values can be assessed by comparing aspects of local crack tip fields, as computed directly from the analysis, with fields deduced from equations (1 a,b) and the J-distributions given by Fig. 2. Figure 3 shows normal stress, \( \sigma_{zz} \), on the plane ahead of the crack front at deepest penetration, normalized by \( \sigma_{hr} \), the corresponding stress component as given by equation (1a), plotted vs. normalized distance, \( r/(J/\sigma_0) \), at normalized load levels of \( \sigma_{zz}/\sigma_0 = 0.47, 0.67, \) and 0.97.

At small distances, the three curves tend asymptotically to unity as they must. However, the steepness of the approach increases markedly at higher loads. At the highest load, close agreement is obtained only at normalized distances well within the finite blunting zone, thus precluding HRR dominance. An arbitrary, but plausible
criterion for dominance is that actual stress in a small geometry change crack analysis should be within 10% of the HRR values at all points closer than 2.5 \( J/\sigma_0 = 5\delta_e \) from the tip. By this criterion the lowest load level shown is clearly \( J \)-dominated, the intermediate load is borderline regarding dominance, while the highest load level is well beyond dominance. The limit of the stringent 10%/2.5(\( J/\sigma_0 \)) criterion occurs in this problem at \( \sigma_{zz}/\sigma_0 = 0.55 \), and, using Fig. 2, this corresponds to a critical value of \( \mu_{cr} = (t-a)/(J/\sigma_0) \geq 500 \).

![Graph showing J-Integral evolution with load magnitude at various positions, \( \phi \) along crack front.](image)

Figure 2. J-Integral evolution with load magnitude at various positions, \( \phi \) along crack front.

![Graph showing normal stress vs. distance at centerline on plane ahead of semi-circular surface crack front at small, intermediate, and full-scale yielding.](image)

Figure 3. Normal stress vs. distance at centerline on plane ahead of semi-circular surface crack front at small, intermediate, and full-scale yielding.
Detailed 3D continuum nonlinear solutions of the sort described in the previous section will continue to be needed in order to better understand limits of dominant singularity approaches to the ductile fracture of structures. However, the great costs of both computer time and data preparation and reduction associated with such solutions will greatly limit their scope, so the development of simplified, but reasonably accurate models is of great concern.

For the broad class of problems dealing with part-through surface-cracks in plates and shells, the so-called "line-spring" model [14-16] has become a useful such compromise. Recently Shawki et. al., [17] have made further developments to this model. The basic features of the line-spring model are relatively well-known [14-16] and will not be elaborated on here. Fig. 4 shows how a part-through surface crack is modeled as a through crack having a generalized foundation supporting generalized loads \( Q(x) \) with work-conjugate displacements \( q(x) \). The stiffness connecting \( q \) and \( Q \) varies with local depth \( a(x) \), and is typically derived from the extra "cracked" compliance of a (SEN) specimen of crack depth \( a(x) \), thickness \( t \), and subject to \( Q \).

![Figure 4a. Schematic section of a surface crack with varying depth \( a(x) \) and projected length \( 2c \) in a plate of thickness \( t \).](image1)

![Figure 4b. Line-Spring model of equivalent through-crack with generalized foundation across model through crack-faces.](image2)
Models for the SEN behavior were previously available for linear elasticity and for flow, or incremental theory plasticity [15, 16], but had the undesirable feature of an abrupt transition, of rather low accuracy, connecting the small-scale, and fully plastic yielding regimes. Shawki et al. smooth this transition region, improving the model accuracy in two ways. (It should be noted that this is just the region where J may be expected to be of use in ductile structures). First, an effective crack length concept, similar to that used in moderate scale yielding LEFM, was introduced.

Secondly, additive power law generalized displacements \( q \alpha Q^n \) were also introduced. Since an accurate calibration of the power law SEN specimen subject to arbitrary combinations of tension and bending is lacking, Shawki et al. developed an approximate calibration of deeply-cracked SEN specimens based on recent solutions [18] of a semi-infinite crack approaching the boundary of a half space, leaving a remaining ligament subject to tension and bending. Nakamura and Shawki [19] showed that that the approximate model accurately reproduces solutions [20, 21] of the finite SEN specimen providing crack depth to thickness exceeds \( a/t \approx .375 \).

Figure 5 shows normalized values of J vs. load T at deepest penetration for a tensile-loaded plate containing a semi-elliptical surface flaw of relative depth \( a/t \approx 0.6 \) and aspect ratio \( a/c \approx 0.24 \). Here \( T_0 = \sigma_0 (bt-\pi ac/2) \) is a reference load. Both 3D continuum and line-spring solutions are shown for two hardening exponents, \( n = 3 \) and 10. For each material model, the Ramberg Osgood parameter \( \alpha = 1 \). As can be seen, the agreement is remarkable. Fig. 6 compares the variation of J along the crack front at various load levels for \( n = 10 \). Here \( \phi = \cos^{-1}(-1-x^2/c^2) \) is a parametric angle widely used for representing points along elliptical crack fronts, with \( \phi = 0 \) at deepest penetration and \( \phi = \pi/2 \) at the intersection with the free surface. Again, agreement is generally good along most of the crack front at all load levels. It should be noted that the line-spring solutions in Figs. 5, 6 required roughly one order of magnitude less computing time than the corresponding continuum solution.

![Figure 5. Normalized centerline J vs. tensile load T for a plate with a surface crack. Continuum and line-spring solutions are shown for material models with \( n = 3 \) and 10.](image-url)
Finally, Shawki et. al. have also noted that the ratio of tension to bending obtained from line-spring solutions can be used with prior study of HRR dominance [9] to obtain approximate assessments of loss such dominance without resort to detailed continuum analysis.

ACKNOWLEDGMENTS

This work was supported by D.O.E. under grant DE-FC02-85ER13331 with MIT and the Idaho National Engineering Laboratory. Computations were performed on a Data General MV10000 computer donated to M.I.T. by Data General Computer, and on an Alliant FX/8 computer through sponsorship of DARPA. The contributions of co-workers T. Nakamura, M. K. Delatte, Y. Wang, and T. Shawki are appreciated.

REFERENCES


SOLIDS WITH MULTIPLE CRACKS AND SOME RELATED PROBLEMS

Mark Kachanov
Department of Mechanical Engineering
Tufts University, Medford, MA 02155

ABSTRACT

A simple method of analysis of crack interactions in elastic solids based on a superposition technique is presented. The method applies to both two- and three-dimensional crack arrays of arbitrary geometry and yields approximate analytical solutions accurate up to quite close distances between cracks. Simple expressions for the effective elastic properties of a solid with any deterministic field of cracks are obtained; these expressions are asymptotically exact in the case of weak interactions and can be expected, on the other hand, to yield accurate results at high crack densities.

INTRODUCTION

We consider an elastic solid with \( N \) cracks \( \sigma \)-jected to remote loading \( \sigma^\infty \). This problem is equivalent to the one with tractions \( \nabla \cdot \sigma^\infty \) applied to the crack faces (\( n_i \) is a unit normal to the \( i \)-th crack) and stresses vanishing at infinity. The latter problem can be represented as a superposition of \( N \) problems, each involving only one crack but loaded by unknown tractions (induced, on a given crack line, by the other cracks and by the remote loading). According to a conventional approach to the problem, these tractions are interrelated through a system of integral equations; the latter are usually solved by polynomial approximations of various types. A new method of analysis is briefly outlined below (see, also, [1]). It is based on the idea of self-consistency applied to the average tractions on individual cracks. The accuracy generally depends on the density of cracks and remains good up to quite close distances between cracks. The method is first demonstrated on a "test" problem with known exact solution and then formulated for the general case.

TEST PROBLEM: TWO COLLINEAR CRACKS

The problem of two collinear cracks of equal length in an infinite plate loaded by uniform remote traction \( \sigma^\infty \) has an excellent analytical solution [2] and can, therefore, serve as a test for the proposed method.

The problem is equivalent to the following one: normal traction \( \sigma^\infty \) is applied to the crack faces and stresses vanish at infinity. The latter problem can be represented as a superposition of two problems, each involving only one isolated crack but loaded by unknown tractions (see Figure 1 where the parameter \( k \) characterizes the ratio of the distance \( 2k \) between cracks to the crack length \( 1-k \)). The traction on crack 2, \( (k, 1) \) is

\[
p_2(x) = \sigma^\infty + p'(x)
\]

where \( p'(x) \) denotes the stress \( \sigma_{yy} \) generated along the line \( (k,1) \) by crack 1, the
the latter being loaded by

$$p_1(x) = p^\infty + p''(x)$$  \hspace{1cm} (2)

where \(p''(x)\) is the crack 2-generated stress. (Interrelating \(p'\) and \(p''\) through integral equations would constitute a conventional approach to the problem).

Represent \(p_1(x)\) as a sum of its average \(p_1 = \langle p^\infty + p'' \rangle = p^\infty + \langle p'' \rangle\) and the difference \(p'' - \langle p'' \rangle\) having zero average (Fig. 2). The key assumption of the method is to neglect the traction on crack 2 due to the load \(p'' - \langle p'' \rangle\) on crack 1. Thus the traction on a given crack induced by the other crack is taken as response of the latter to the uniform average traction on it; the impact on crack 2 of the traction nonuniformities with a zero average on crack 1 is neglected. This assumption results in a major simplification of the problem. Indeed, the traction \(p'\) induced on crack 2 by

\[
\begin{align*}
\langle p^\infty + p'' \rangle & = \langle p^\infty \rangle + \langle p'' \rangle \\
\langle p'' \rangle & = \langle p'' \rangle \\
\langle p'' \rangle & = \langle p'' \rangle 
\end{align*}
\]

Figure 2. Representation of traction as a sum of its average and a non-uniformity.

The factor \(A\) characterizes attenuation of the average traction in transmission of stress from crack 1 onto the crack 2 line. Note that \(A < 1\) and decreases from \(\sqrt{2} - 1\) to
0 when k changes from 0 to 1. Obviously, A is the same in transmissions from crack 1 onto the crack 2 line and vice versa. 

Now \( <p_1> \) is readily found in terms of \( \Lambda \): taking the average of (1)

\[
<p_2> = p^\infty + <p'> = p^\infty + \Lambda <p_1>
\]

together with the condition \( <p_1> = <p_2> = <p> \) (symmetry of the configuration) yields

\[
<p> = \frac{p^\infty}{1 - \Lambda}
\]

(5)

Formula (5) shows increase of the average traction due to the crack interaction. Using the general formula (10) for the SIFs at the tips of a crack loaded by a given traction, the latter being taken from (3), one obtains the following expressions for the SIFs at the outer and inner tips, correspondingly:

\[
K_I (1) = K_I^O \left\{ 1 + \frac{1}{1 - \Lambda} \frac{1}{\pi (1-k)} \left[ 2 \xi - k(k+1)K - \frac{\pi}{2} (1-k) \right] \right\}
\]

(6)

\[
K_I (k) = K_I^O \left\{ 1 + \frac{1}{1 - \Lambda} \frac{1}{\pi (1-k)} \left[ -2 \xi + (k+1)K - \frac{\pi}{2} (1-k) \right] \right\}
\]

where \( K_I^O = p^\infty \sqrt{\pi (1-k)}/2 \) is the SIF for an isolated crack and \( K, \xi \) are complete elliptic integrals of the argument \( k' = \sqrt{1 - k^2} \) of the first and second kind, respectively. This completes the solution.

These results are compared with the exact ones in Table 1. The agreement is very good. The error becomes noticeable for closely located cracks but even at \( k = .05 \) (distance between the cracks is one order of magnitude smaller than the crack length), the error is only .4% for \( K_I (k) \) and .2% for \( K_I (1) \).

<table>
<thead>
<tr>
<th>k</th>
<th>( K_I (k)/K_I^O ) (inner tip)</th>
<th>Error</th>
<th>( K_I (1)/K_I^O ) (outer tip)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2</td>
<td>1.112</td>
<td>1.112</td>
<td>.3%</td>
<td>1.052</td>
</tr>
<tr>
<td>.1</td>
<td>1.255</td>
<td>1.251</td>
<td>.32</td>
<td>1.086</td>
</tr>
<tr>
<td>.05</td>
<td>1.473</td>
<td>1.452</td>
<td>1.42</td>
<td>1.120</td>
</tr>
<tr>
<td>.02</td>
<td>1.905</td>
<td>1.809</td>
<td>5.02</td>
<td>1.159</td>
</tr>
<tr>
<td>.01</td>
<td>2.372</td>
<td>2.134</td>
<td>10%</td>
<td>1.184</td>
</tr>
</tbody>
</table>

### ARBITRARY CRACK ARRAY

Application of the method to an arbitrary 2-D array of N cracks \( \xi_1, \ldots, \xi_N \) with unit normals \( n_i \) and remove loading \( \sigma^\infty \) is straightforward. The equivalent problem - with the tractions \( n_i \sigma^\infty \) applied to the i-th crack faces and stresses vanishing at infinity - can be reduced, by superposition, to N problems, each containing one isolated crack loaded by unknown tractions. Problem 1, for instance, will contain the...
crack \ell_1 \text{ loaded by } n_1 \sigma^\infty \text{ plus the sum of the tractions generated along the } \ell_1\text{-line by each of the cracks } \ell_2^\perp \text{ (treated as isolated cracks in the otherwise continuous material).}

According to the key idea of the method, we represent the latter tractions as responses of the cracks \ell_2^\perp to the (yet unknown) uniform loads–traction averages on them. Modes I and II will be interrelated: normal (shear) loading of a crack will generate both normal and shear tractions along the other crack lines. Thus, denoting by \( p_1 \) and \( \tau_1 \) the normal and shear tractions on \( \ell_1 \) and by \( \left< p_1 \right> \) and \( \left< \tau_1 \right> \) – their averages, we have the following tractions on crack \( \ell_1 \) (\( \ell_1 \) is a current point on \( \ell_1 \)):

\[
\begin{align*}
p_1(\xi) &= p_1^{\infty} + n_1 \cdot \left[ \sum_{2}^{N} \sigma_{2}^{\ell_1}(\xi) p_2 + \tau_{2}^{\ell_1}(\xi) \tau_2 + \ldots + \sigma_{N}^{\ell_1}(\xi) p_N + \tau_{N}^{\ell_1}(\xi) \tau_N \right] \cdot n_1 \\
\tau_1(\xi) &= \tau_1^{\infty} + n_1 \cdot \left[ \sum_{2}^{N} \sigma_{2}^{\ell_1}(\xi) p_2 + \tau_{2}^{\ell_1}(\xi) \tau_2 + \ldots + \sigma_{N}^{\ell_1}(\xi) p_N + \tau_{N}^{\ell_1}(\xi) \tau_N \right] \cdot (I - n_1 n_1)
\end{align*}
\]

where \( p_1^{\infty} = n_1 \cdot \sigma^\infty \cdot n_1 \) and \( \tau_1^{\infty} = n_1 \cdot \sigma^\infty \cdot (I - n_1 n_1) \) are the normal and shear tractions induced on \( \ell_1 \) by the remote loading (\( I \) and \( n_1 n_1 \) denote a unit tensor and dyadic product of two vectors \( n_1 \)); \( \sigma_i^\ell_1, \tau_i^\ell_1 \) are the stress fields generated by the \( i \)-th crack loaded by uniform tractions (normal and shear, correspondingly) of unit intensity; they are "standard" fields given by elementary functions (see, for example, [2]). The quantity in the brackets is the stress tensor induced along \( \ell_1 \) by the other cracks loaded by the tractions averages on them.

Averaging (7) along \( \ell_1 \) yields

\[
\begin{align*}
\left< p_1 \right> &= p_1^{\infty} + \Lambda_{121}^{nn} p_2 + \Lambda_{121}^{nT} \tau_2 + \ldots + \Lambda_{1N1}^{nn} p_N + \Lambda_{1N1}^{nT} \tau_N \\
\left< \tau_1 \right> &= \tau_1^{\infty} + \Lambda_{121}^{TT} \tau_2 + \Lambda_{121}^{TT} \tau_2 + \ldots + \Lambda_{1N1}^{TT} \tau_N + \Lambda_{1N1}^{TT} \tau_N
\end{align*}
\]

where the transmission \( \Lambda \)-factors characterize transmissions of the average normal and shear tractions; for example, \( \Lambda_{13}^{TT} \) is the average shear traction on crack \( \ell_3 \) resulting from the normal uniform load of unit intensity on crack \( \ell_1 \).

Equations (8) and similar equations on \( \ell_2, \ldots, \ell_N \) constitute a system of \( 2N \) linear algebraic equations ("self-consistency" equations) for the average tractions \( \left< p_k \right>, \left< \tau_k \right> \). Denoting by \( \left< t_k \right> = \left< p_k \right>, \left< \tau_k \right> \) the average traction vector on the \( k \)-th crack, we can write this system in a form of \( N \) vectorial linear algebraic equations

\[
\begin{align*}
\left< t_k \right> = t_k^{\infty} + \Lambda_{ik}^{nT} \cdot \left< t_i \right> \quad \text{(sum over } i = 1, \ldots, N; i \neq k) \\
\left< t_k \right> = t_k^{\infty} + \Lambda_{ik}^{TT} \cdot \left< t_i \right>
\end{align*}
\]

where the tensorial element \( \Lambda_{ik} \) gives the average traction vector generated along the \( k \)-th crack line by an (isolated) \( i \)-th crack loaded by a uniform traction of arbitrary direction and unit intensity. It is given by the expression

\[
\Lambda_{ik} = n_k \cdot \sigma_{i}^n + \sigma_{i}^T \cdot n_k
\]

involving averaging of the "standard" stress fields \( \sigma_i^n, \sigma_i^T \) generated by the \( i \)-th crack over \( \ell_k \) (\( \ell_k \) is a unit vector tangent to \( \ell_k \) and \( n_k \cdot \sigma_i^n, \sigma_i^T \cdot n_k \) are dyadic (tensor products). Noting that the diagonal elements \( \Lambda_{ii} \), \( \Lambda_{22}, \ldots \) characterizing the "interaction of a crack with itself" are unit tensors \( I \), we can rewrite (9a) in a compact form

\[
\left( \Lambda - \delta_{ik} I \right) \cdot \left< t_i \right> = -t_k^{\infty}
\]
where the conventional summation agreement over all \( i = 1, \ldots, N \) is observed.

Thus, crack interactions are described, in the framework of our method, by the interaction matrix \([A_{ij}]\). It is, generally, nonsymmetric: \( A_{ij} \neq A_{ji} \) (for instance, the impact of a large crack on a small one is larger than vice versa).

Note that the interaction matrix \([A_{ij}]\) provides an intrinsic description of the crack array, reflecting its geometry but independent of the remote loading conditions; the latter affect only the right-hand parts of (3.3b).

After the average tractions <\( \tau_i \)> are determined from this system, tractions \( p_i(\xi), \tau_i(\xi) \) on cracks are found from (7) and similar equations for \( i = 2, \ldots, N \); the SIFS are readily obtained from

\[
K_i(\frac{\xi}{\lambda}) = \frac{1}{\sqrt{\pi \xi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\xi^2 + \xi'}} \left\{ \frac{p_i(\xi)}{\tau_i(\xi)} \right\} d\xi
\]

The mode III loading can be analyzed along the same lines. Since it does not interact with modes I and II, the mode III analysis can be done separately.

In the 3-D case, the method remains essentially unchanged; the "standard" stress fields may be given by special, rather than elementary functions and the SIFS (generally variable along the crack edges) are to be found by the 3-D analogues of integrations (10). The method can, therefore, be applied to arbitrary arrays of elliptical cracks (for which the "standard" field generated by a uniform loading on a crack is known).

Thus, the method involves: (1) finding the transmission factors, by averaging the "standard" stress fields generated by uniformly loaded cracks along the lines of the other cracks, (2) solving a system of linear algebraic equations for the average tractions, and (3) finding the SIFS by integrations of the type (10).

**WEAK CRACK INTERACTIONS**

The test problem considered above shows that the errors in SIFS vanish as the distance between cracks increases and interactions become weak. This fact is general: the method is asymptotically exact for weak crack interactions. We call the interactions weak if the average tractions <\( \lambda p_i \), <\( \lambda \tau_i \)> induced on each crack by the other cracks are small as compared to the \( \sigma^\infty \)-induced \( \sigma_i^\infty \), \( \tau_i^\infty \). Then all the transmission factors \( \Lambda << 1 \) and, to within small values of higher order, the traction on a given crack is a sum of \( \sigma_i^\infty \) and the tractions induced on it by the other cracks, the latter being embedded in the \( \sigma^\infty \)-field. (This follows as a first order solution of the system (7) obtained by one iteration, with \( p_i(\xi) = p_i^\infty \), \( \tau_i(\xi) = \tau_i^\infty \) being an obvious zeroth iteration). Thus, instead of

\[
t_k(\xi) = t_{k}^\infty + n_k \cdot (<p_i^\infty \sigma_{i1} + <\tau_i^\infty \sigma_{i1}>)
\]

(vectorial form of (7), we have

\[
t_k(\xi) = t_{k}^\infty + n_k \cdot (p_{i1}^\infty \sigma_{i1} + \tau_{i1}^\infty \sigma_{i1}^\top)
\]

so that finding traction averages becomes unnecessary.

Applicability of the weak interactions approximation depends on whether the inequalities for tractions (\( <\lambda p_i, <\lambda \tau_i> << (\sigma_p^\infty, \tau_\infty^\infty) \) hold; they are much less restrictive than the geometrical condition (distances between cracks \( \gg \) (crack sizes) and, depending on the geometry of the crack array, may be satisfied for relatively closely
REMOTELY LOCATED CRACKS

As noted above, the condition of weak interactions \(<\Delta p>\), \(<\Delta \tau>\) \(\ll (p^\infty, \tau^\infty)\) does not impose any explicit conditions on the geometry of the crack array and is much less restrictive than the assumption of remotely located cracks (distances between cracks \(\gg\) crack sizes). If, however, the latter assumption is made, further simplifications result and the solutions become quite elementary.

Since gradients of the stresses generated by cracks attenuate faster than the stresses themselves, the traction induced on a given crack line by the other cracks can be taken approximately constant. This constant can be taken, to within small values of a higher order, as the traction \(n_i \cdot \nabla_k \sigma_k(0_i)\) evaluated at the center \(0_i\) of the given crack \(l_i\), where the "standard" fields \(\sigma_k\) can be substituted by their remote asymptotics.

Therefore, both averaging of the "standard" fields and solving a system of linear algebraic equations for averages become unnecessary and the entire procedure is reduced to evaluation of the remote "standard" fields at the other cracks' centers.

The approximation of remotely located cracks, as defined above, should be distinguished from the so-called "small concentration" approximation (terminology often used in models for the effective elastic properties of a medium with cracks). The latter treats each crack as embedded into the remotely applied field \(\sigma^\infty\); stress fields generated by the cracks are simply summed up, and crack interactions are entirely neglected. The approximation of remotely located cracks, on the other hand, provides a first order correction due to crack interactions.

Note that this first order correction depends not only on the density of cracks and their orientations but on their mutual locations (statistics of the crack centers) as well. The latter can be disregarded in the approximation of noninteracting cracks (infinitesimal crack density) only. This remark is relevant for the problem of effective elastic properties of cracked solids at finite crack densities.

ON THE PROBLEM OF EFFECTIVE ELASTIC PROPERTIES OF SOLIDS WITH CRACKS

Below, we outline how to apply the presented method to the problem of effective elastic properties of solids with cracks.

The problem is formulated as finding a fourth rank tensor \(\mathbf{C}^{\text{eff}}\) relating the average over a representative volume \(V\) strain \(<\varepsilon>\) to the remotely applied stress \(\sigma^\infty\): \[<\varepsilon> = \mathbf{C}^{\text{eff}} : \sigma^\infty\] (Semicolon denotes contraction over two indices.) The commonly used starting point is a representation of \(<\varepsilon>\) in the form:

\[<\varepsilon> = \mathbf{C}^0 : \sigma^\infty + \frac{1}{2V} \int_{S_i} (n_b + b_n) \, dS\] (13)

Notations are as follows: \(\mathbf{C}^0\) is a tensor of elastic compliance of the material without cracks, \(n_i\) is a unit normal to the \(i\)-th crack with a surface \(S_i\); \(b_i\) is the displacement discontinuity across \(S_i\) (relative displacement of the crack faces). For flat cracks, \(n = \text{const}\) along each of the cracks so that

\[<\varepsilon> = \mathbf{C}^0 : \sigma^\infty + \frac{1}{2V} \int_{S_i} (n \langle b_i \rangle + \langle b \rangle n) \, dS_i\] (14)

where \(b_i\) denotes the average of \(b_i\) over \(S_i\). In the 2-D case, surfaces \(S_i\) are to be
substituted by lines \( \lambda_1 \) and representative volume \( V \) by representation area \( A \).

Thus, the problem is reduced to the determination of the average displacement discontinuities \( \langle b_1 \rangle \) on cracks; if we could relate them to \( \sigma^\infty \) and to the crack array geometry, the effective elastic properties \( \varepsilon_{\text{eff}} \) would have been found.

The existing models, aside from the approximation of noninteracting cracks (small crack density) belong to one of the modifications of the self-consistent or differential schemes. Limitations and certain conceptual difficulties of these models were discussed in a recent survey by Hashin [3]. Note, also, that it appears difficult to incorporate the information on mutual locations of cracks (statistics of crack centers) into these schemes.

A new approach to the problem, based on the presented method can be suggested. The problem that arises is to relate the average displacement discontinuities \( \langle b_1 \rangle \) in (14) to the quantities \( K(\pm\lambda_1) \) and \( \langle t_1 \rangle \) (SIFs and average tractions on cracks) provided by the method.

Although the vectors \( \langle b_1 \rangle \) and \( \langle t_1 \rangle \) are not exactly proportional (otherwise a traction distribution with a zero average would have always generated a COD with a zero average), it appears that their proportionality can be assumed with a high degree of accuracy (see [1]):

\[
\langle b \rangle = \frac{\pi}{E} \langle t \rangle
\]

(15)

In the 3-D configurations, the vectors \( \langle b \rangle \) and \( \langle t \rangle \) are not parallel (although the angle between them is relatively small for the penny-shaped cracks) and, therefore, relation of the type (15) is to be formulated separately for the normal and shear components \( \langle b_n \rangle \sim \langle t_n \rangle \) and \( \langle b_\tau \rangle \sim \langle t_\tau \rangle \).

Formula (15) means that the contribution of the traction nonuniformities having zero average into the average COD can be neglected. Equation (15) implies, also, that calculation of SIFs becomes unnecessary and the problem of effective elastic properties is reduced simply to inversion of the interaction matrix. From the physical point of view, relation (15) means that the impact of cracks on the effective properties is "felt" only through the transmission factors characterizing attenuation of the traction averages. Since (15) has such farreaching consequences, it requires a further examination, for various crack configurations.

Since the presented method of analysis of crack interactions is accurate at close distances between cracks, the results for the effective elastic moduli can be expected to be applicable at high crack densities. Locations of the crack centers will be taken into account through the interaction matrix. The problem requiring further attention is incorporation of the statistical information on the crack array into the model.

ACKNOWLEDGMENT

Partial support of the U.S. Department of Energy is gratefully acknowledged.

REFERENCES


LIST OF PARTICIPANTS

J. D. Achenbach  
Civil Engineering and  
Applied Mathematics  
Northwestern University  
Evanston, IL  60201

Hassan Aref  
Department of AMES  
University of California-San Diego  
La Jolla, CA  92093

George Bankoff  
Department of Chemical Engineering  
Technological Institute  
Northwestern University  
Evanston, IL  60201

Thomas J. Bruno  
National Bureau of Standards  
325 Broadway  
Boulder, CO  80303

In-Hwa Chang  
Massachusetts Institute of Technology  
77 Massachusetts Avenue  
Cambridge, MA  02139

Jacob N. Chung  
Mechanical and Materials Engineering  
Washington State University  
Pullman, WA  99164

Ara Chutjian  
Jet Propulsion Laboratory  
4800 Oak Grove Drive  
Pasadena, CA  91109

Samuel K. Clark  
315 Auto Laboratory  
University of Michigan  
Ann Arbor, MI  48109

John F. Elliott  
Materials Science and Engineering  
Massachusetts Institute of Technology  
77 Massachusetts Avenue  
Cambridge, MA  02139

Jeffrey Fanton  
Ginzton Laboratory  
Stanford University  
Stanford, CA  94305

Michael B. Frish  
Physical Sciences, Inc.  
P. O. Box 3100  
Andover, MA  01810

William Ross Hamel  
Oak Ridge National Laboratory  
P. O. Box X  
Oak Ridge, TN  37831

Lambertus Hesselink  
Department of Aero/Astronautics  
Stanford University  
113 Durand Building  
Stanford, CA  94305

David K. Hsu  
Ames Laboratory  
Iowa State University  
Ames, IA  50011

Yukikazu Iwasa  
Massachusetts Institute of Technology  
77 Massachusetts Avenue  
Cambridge, MA  02139

John A. Johnson  
EGSG Idaho, Inc.  
P. O. Box 1625  
Idaho Falls, ID  83415

Daniel D. Joseph  
Department of Aerospace Engineering  
and Mechanics  
University of Minnesota  
110 Union Street  
Minneapolis, MN  55455

Mark Kachanov  
Department of Mechanical Engineering  
Tufts University  
Medford, MA  02155

Yong-Seog Kim  
Massachusetts Institute of Technology  
77 Massachusetts Avenue  
Cambridge, MA  02139

Max Klein  
Gas Research Institute  
8600 W. Bryn Mawr  
Chicago, IL  60631

Erhard Krempel  
Department of Mechanical Engineering,  
Aeronautical Engineering and Mechanics  
Rensselaer Polytechnic Institute  
Troy, NY  12180

Johanna M. H. Levei-Sengers  
National Bureau of Standards  
Division 774  
Gaithersburg, MD  20899