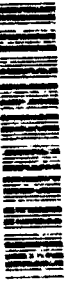


LOAN DOCUMENT

PHOTOGRAPH THIS SHEET

AD-A261 994



DTIC ACCESSION NUMBER

LEVEL

INVENTORY

AFOSR-TR-93-0116

DOCUMENT IDENTIFICATION

Dec 92

DISTRIBUTION STATEMENT

Approved for public release:
Distribution Unlimited

DISTRIBUTION STATEMENT

ACCESSION FOR

NTIS ☐ GRA&I ☒

DTIC ☐ TRAC ☐

UNANNOUNCED ☐

JUSTIFICATION ☐

BY

DISTRIBUTION/

AVAILABILITY CODES

DISTRIBUTION

AVAILABILITY AND/OR SPECIAL

A-1

DISTRIBUTION STAMP

DTIC
ELECTE
MAR 10 1993
S C D

DATE ACCESSIONED

DATE RETURNED

98 3 4 072

98 3 4 013

DATE RECEIVED IN DTIC

93-04700



REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

H
A
N
D
L
E
W
I
T
H
C
A
R
E

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1992
SUMMER FACULTY RESEARCH PROGRAM
(SFRP) REPORTS

VOLUME 6

ARNOLD ENGINEERING DEVELOPMENT CENTER
CIVIL ENGINEERING LABORATORY
FRANK J. SEILER RESEARCH LABORATORY
WILFORD HALL MEDICAL CENTER

RESEARCH & DEVELOPMENT LABORATORIES

5800 UPLANDER WAY
CULVER CITY, CA 90230-6608

SUBMITTED TO:

LT. COL. CLAUDE CAVENDER
PROGRAM MANAGER

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE

WASHINGTON, D.C.

DECEMBER 1992

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank) 2. REPORT DATE
28 Dec 92 Annual 1 Sep 91 - 31 Aug 92

3. TITLE AND SUBTITLE

1992 Summer Faculty Research Program (SFRP)
Volumes 1 - 16

V-6

F49620-90-C-0076

Mr Gary Moore

4. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

AFOSR-TR- 83 0116

Research & Development Laboratories (RDL)
5800 Uplander Way
Culver City CA 90230-6600

5. MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NI
110 Duncan Ave., Suite B115
Bldg 410
Bolling AFB DC 20332-0001
Lt Col Claude Cavender

6. SUPPLEMENTARY NOTES

7. DISTRIBUTION AVAILABILITY STATEMENT

UNLIMITED

8. ABSTRACT (Maximum 200 words)

The purpose of this program is to develop the basis for continuing research of interest to the Air Force at the institution of the faculty member; to stimulate continuing relations among faculty members and professional peers in the Air Force to enhance the research interests and capabilities of scientific and engineering educators; and to provide follow-on funding for research of particular promise that was started at an Air Force laboratory under the Summer Faculty Research Program.

During the summer of 1992 185 university faculty conducted research at Air Force laboratories for a period of 10 weeks. Each participant provided a report of their research, and these reports are consolidated into this annual report.

9. SUBJECT TERMS

10. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	11. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	12. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	UL

13. DISTRIBUTION STATEMENT

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1992
SUMMER FACULTY RESEARCH PROGRAM (SFRP) REPORTS

VOLUME 6

ARNOLD ENGINEERING DEVELOPMENT CENTER
CIVIL ENGINEERING LABORATORY
FRANK J. SEILER RESEARCH LABORATORY
WILFORD HALL MEDICAL CENTER

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Lt. Col. Claude Cavender

Program Manager, RDL
Billy Kelley

Program Administrator, RDL
Gwendolyn Smith

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1992

PREFACE

This volume is part of a 16-volume set that summarizes the research accomplishments of faculty, graduate student, and high school participants in the 1992 AFOSR Summer Research Program. The current volume, Volume 6 of 16, presents the final research reports of faculty (SFRP) participants at Arnold Engineering Development Center, Civil Engineering Laboratory, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3. Research reports in the 16-volume set are organized as follows:

VOLUME	TITLE
1	Program Management Report
2	Summer Faculty Research Program Reports: Armstrong Laboratory
3	Summer Faculty Research Program Reports: Phillips Laboratory
4	Summer Faculty Research Program Reports: Rome Laboratory
5A	Summer Faculty Research Program Reports: Wright Laboratory (part one)
5B	Summer Faculty Research Program Reports: Wright Laboratory (part two)
6	Summer Faculty Research Program Reports: Arnold Engineering Development Center; Civil Engineering Laboratory; Frank J. Seiler Research Laboratory; Wilford Hall Medical Center
7	Graduate Student Research Program Reports: Armstrong Laboratory
8	Graduate Student Research Program Reports: Phillips Laboratory
9	Graduate Student Research Program Reports: Rome Laboratory
10	Graduate Student Research Program Reports: Wright Laboratory
11	Graduate Student Research Program Reports: Arnold Engineering Development Center; Civil Engineering Laboratory; Frank J. Seiler Research Laboratory; Wilford Hall Medical Center
12	High School Apprenticeship Program Reports: Armstrong Laboratory
13	High School Apprenticeship Program Reports: Phillips Laboratory
14	High School Apprenticeship Program Reports: Rome Laboratory
15	High School Apprenticeship Program Reports: Wright Laboratory
16	High School Apprenticeship Program Reports: Arnold Engineering Development Center; Civil Engineering Laboratory

1992 FACULTY RESEARCH REPORTS

Arnold Engineering Development Center
Civil Engineering Laboratory
Frank J. Seiler Research Laboratory
Wilford Hall Medical Center

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
<u>Arnold Engineering Development Center</u>		
1	Experiences using Model-Based Techniques for the Development of a Large Parallel Instrumentation System	Dr. Ben Abbott
2	Same as above	Dr. Csaba Biegl
3	Data Reduction of Laser Induced Fluorescence in Rocket Motor Exhausts	Dr. Stephen H. Cobb
4	Feasibility of Wavelet Analysis for Plume Data Study	Dr. Kenneth R. Kimble
5	Current Flow in the Plasma Erosion Opening Switch	Dr. Carlyle E. Moore
6	Extension of the SMIRF Flare Model to Complex Geometrical Shapes	Dr. Olin Perry Norton
7	Fourier Transform Spectrometric Infrared Detection of Airborne Solvents	Dr. Randolph S. Peterson
8	Computational Methods for Calculating Radiation Transport from Pulsed X-Ray Sources	Dr. Richard M. Roberds
<u>Civil Engineering Laboratory</u>		
9	Characterization of Seagrass Meadows in St. Andrew (Crooked Island) Sound, Northern Gulf of Mexico: Preliminary Findings	Dr. Sneed B. Collard
10	A Preliminary Study of the Weathering of Jet Fuels in Soil Monitored by SFE with GC Analysis	Dr. Larry E. Gerdorn
11	Preliminary Numerical Model of Groundwater Flow at the MADE2 Site	Dr. Donald D. Gray
12	Decision Analysis for Selection of Halon 1301 Replacement Agents	Dr. Charles J. Kibert
13	Hydraulic Conductivity Variability, Kriging, Trend Surfaces and Travelling Waves with Nonlinear, Nonequilibrium Adsorption	Dr. Valipuram S. Manoranjan
14	Immobilized Cell Bioreactor for 2,4-Dinitrotoluene Degradation	Dr. Kenneth F. Reardon
15	Uniaxial Stress-Strain Behavior of Unsaturated Soils at High Strain Rates	Dr. George E. Veyera
16	Application of Fiber-Optic Laser Fluorescence Spectroscopy to Environmental Monitoring	Dr. Brian S. Vogt

<u>Report Number</u>	<u>Report Title</u>	<u>Author</u>
<u><i>Civil Engineering Laboratory (cont'd)</i></u>		
17	Applications of SMES in Air Force	Dr. Xingwu Wang
<u><i>Frank J. Seiler Research Laboratory</i></u>		
18	Multivariable Transfer Functions and Optimal Passive Damping for a Slewing Piezoelectric Laminate Beam	Dr. Thomas E. Alberts
19	Velocity and Vorticity Measurements in Transient Oscillatory Separating Boundary Layer Flows	Dr. B. Terry Beck
20	NMR Relaxation Studies of Microdynamics in Chloroaluminate Melts	Dr. William R. Carper
21	Characterization of Materials for Non-Linear Optical Thin Films	Dr. Thomas M. Christensen
22	Modeling of the NLO Properties of Silica	Dr. Ravi Pandey
23	Electrochemistry in Lithium Chloride Buffered-Neutral Room Temperature Melts	Dr. Bernard J. Piersma
<u><i>Wilford Hall Medical Center</i></u>		
24	Enhanced Physiologic Monitoring of Closed Head-Injury	Dr. Michael L. Daley
25	Temperature Effects on Aqueous Polymer and Biopolymer Solution Viscosities and on Erythrocyte Sedimentation Rates and Cell Volumes in Mammalian Blood	Dr. W. Drost-Hansen

ARNOLD ENGINEERING DEVELOPMENT CENTER

EXPERIENCES USING MODEL-BASED TECHNIQUES
FOR THE DEVELOPMENT OF A
LARGE PARALLEL INSTRUMENTATION SYSTEM

Ben Abbott
Csaba Biegl
Research Faculty
Theodore A. Bapty
PhD. Candidate
Department of Electrical Engineering

Vanderbilt University
400 24th Ave. S.
Nashville, TN 37235-1824

Final Report for:
Summer Research Program
Arnold Engineering Development Center

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

EXPERIENCES USING MODEL-BASED TECHNIQUES
FOR THE DEVELOPMENT OF A
LARGE PARALLEL INSTRUMENTATION SYSTEM

Ben Abbott
Csaba Biegl
Research Faculty
Theodore A. Bapty
PhD. Candidate
Department of Electrical Engineering
Vanderbilt University

Abstract

Experience using a *model-based* approach to develop an 83 processor parallel instrumentation system for turbine engine aeromechanic stress analysis is described. The approach includes using a graphics based editor to describe the structure of the desired signal flow graph as well as the target hardware architecture. Program synthesis techniques are used to automatically transform these models into an executable system.

INTRODUCTION:

The Vanderbilt University Department of Electrical Engineering Measurement and Computing Systems Group in cooperation with the US Air Force at Arnold Engineering Development Center (AEDC), Sverdrup Technology (also at AEDC), and University of Tennessee Space Institute (UTSI) have developed a computer system for on-line data analysis of turbine engine tests in the altitude test cells of AEDC. This new high-speed computer architecture is based on parallel processing concepts. During actual operation, the current 24 channel system *continuously* delivers 200 MFLOPS. This continuous performance index, measured on the actual application program, exceeds the *peak* performance capabilities of many conventional high performance computers.

The Computer Assisted Dynamic Data Analysis and Monitoring System (CADDMAS) uses a heterogeneous architecture including INMOS transputers for communication and general purpose processing, Zoran and Motorola Digital Signal Processors for signal processing operations, Texas Instruments Graphics Processors for on-line graphical display of calculated data, and an Intel 486 based PC to provide an interactive graphical user interface (GUI). The total processor count of the 24 channel system is 83.

Vanderbilt researchers have been working with AEDC for the past five years. Our role is to develop tools and techniques to manage the software complexity of this large, parallel, real-time instrumentation system. Driving factors concerning the CADDMAS system include:

- High performance computing is required. A 200 channel 2 Giga-Flop **sustained** system is planned.
- Due to the need for interactive monitoring of the processed data, some of the computations being performed must be modified on-the-fly.
- Event/Alarm recognition and response time is critical.
- The system needs to be modular, allowing various components of the hardware to be "plugged" together to build systems of different size and capabilities. The corresponding software must be able to deal with these varying architectures.

In order to address the problems associated with the development of CADDMAS, we have used *model-based* techniques coupled with a real-time macro-dataflow execution envi-

ronment. The principal technique is to automatically generate a macro-dataflow computation from a declarative model describing the desired system. This new technique introduced by the Measurement and Computing Systems Group of Vanderbilt is well suited to a large class of signal processing, instrumentation, and control problems. The primary tools and techniques used here are part of the Multigraph Programming Environment, previously developed by the Measurement and Computing group of Vanderbilt. Using the model-based approach provides several attractive features:

- *Program Synthesis* techniques allow the system complexity to be managed in the model-based format. A simple example of this is a signal processing system. In this case, a model is constructed graphically which looks like a signal flow diagram. At system build time, that model is analyzed and the corresponding real-time dataflow graph is generated on the execution platform. The builder is only loosely coupled to the execution system and therefore is not a bottleneck during execution.
- *Scheduling, Synchronization, and Communication* are implicit in the modeling paradigm. The system is synthesized from executable subroutines written in sequential languages.
- *Dynamic Reconfiguration during execution* of a dataflow graph is supported. The normal scenario for such an episode is that conditions in the currently executing system reach a point where they re-trigger the *program synthesis* as a result of an event (e.g. the user interface has requested a change in the current analysis plots). Since the executing graph is real-time and continuously running, it cannot be stopped, rather, it is modified on-the-fly.

This paper describes our experience with the CADDMAS application system including:

- Background information concerning the Multigraph Programming Environment.
- The CADDMAS system architecture.
- Evaluation of and future plans for the *model-based* approach as a result of the CADDMAS system development.

THE MODEL-BASED APPROACH TO INSTRUMENTATION

Development of real-time instrumentation systems requires the crossover of two distinct engineering fields: signal processing and computer engineering. The high level problem has many signal processing issues, yet the computational complexity of the real-time events and parallel processing synchronization causes numerous low level, computer engineering issues to emerge.

We have chosen to combine numeric and symbolic processing to ease the crossover of the two disciplines [1]. The Multigraph Architecture provides a general framework aiding the construction of model-based engineering systems for real-time, parallel computing environments [2]. The main components of the architecture are shown in Figure 1 and described below:

1. The *Symbolic Model Declarations* contain modeling information encoded in a *problem specific* declarative language. For signal processing systems, these declarations are typically of a form that describes the signal flow graph topology as well as specific parameters of the processing blocks in the diagram. On multiprocessor systems, models of the available hardware resources are also included.
2. The *Model Builder* interprets the symbolic model declarations. It instantiates, initializes, and connects the computation nodes of a macro-dataflow graph. This executable dataflow graph is built as specified by the *Symbolic Model*. In multiprocessor systems, an arc of a graph that crosses processor boundaries automatically sets up inter-processor communication.
3. The *Library of Operations* contains basic sequential programs to be assigned as the computation scripts of the specific nodes of the dataflow graph. For signal processing applications, these are typically numeric routines implementing basic computations (e.g. a FFT or filter). These components are usually implemented in standard numeric languages: C, or Fortran.
4. The *Execution Environment* (also called the Multigraph kernel) provides facilities allowing the model builder to dynamically configure the macro-dataflow graph. As well, the kernel provides a scheduler that controls the graph's execution. The kernels

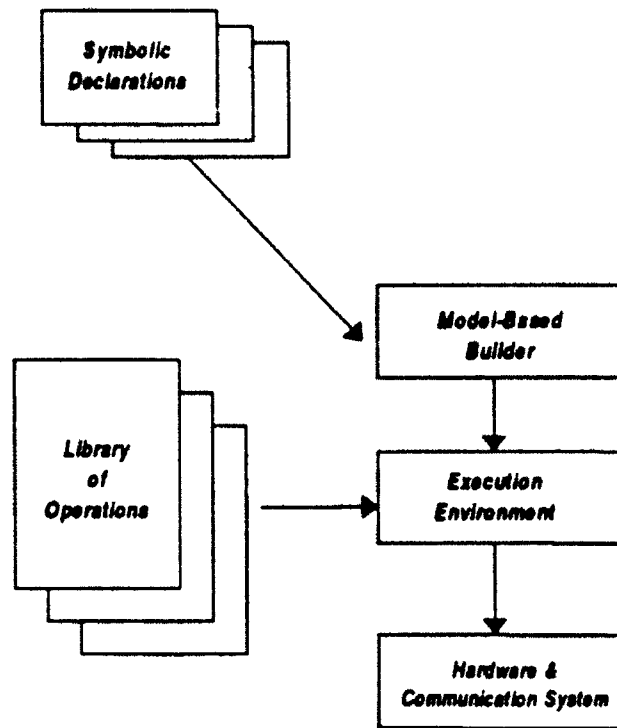


Figure 1: The Multigraph Architecture.

present themselves to the model builder as a single virtual machine even when a set of heterogeneous distributed processing elements are used.

5. The *Hardware and Communication Systems* provide resources to the execution environment needed to run the application as well as a communication path between processors. The Multigraph kernel insulates the user from this level.

EXTENSIONS TO THE DATAFLOW MODEL

The macro-dataflow environment supported by the Multigraph kernel has several unique extensions:

1. *Actornodes* are the computational operators of the dataflow graph. An actornode has a state associated with the kernel scheduler. It may be **inactive**, **active**, **ready**, or **running**. An **inactive** actor will never be chosen for execution by the scheduler. An **active** actor may be chosen but does not have the proper input control tokens to warrant execution at the preset time. A **ready** actor is waiting in a queue for

the processing resources to become available for it to be executed. A **running** actor is currently being executed. Functions to create, destroy, connect input and output ports to datanodes, disconnect from datanodes, activate, deactivated, stop, check the status of, replace, set the context of, set the script of actornodes are all provided for the symbolic interface. An actornode itself is made up from several components:

- (a) The *script* is a piece of code which performs the required operation on data items propagated to the node. It may be coded in almost any language, symbolic or numeric. The script is written in a reentrant fashion so that it can be attached to more than one actor simultaneously. Special Multigraph kernel calls are available to a script programmer for receiving the input control tokens and propagating the resulting token(s). As well, special memory management routines are provided by the kernel for increased performance. Further kernel calls are available for error handling within the kernel.
- (b) The *context* is a static local memory section for a particular actor. It provides two basic services for the actornode. First, it allows the script to have a place to save its state. Thus the actornodes functionality may be changed due to previous values received as input control tokens. Second, it provides an interface to the symbolic builder and control functions such that parameters may be set during the build operation that will affect the operation of the script during execution. For example, a particular signal processing control graph may have an actornode that receives an input vector of time domain data samples, performs a fast fourier transform, and then propagates the result. One implementation of such an actor might be to provide the radix and input data packet size information at build time to the actor through its context. In this case, the context provides a simple interface for the model builder layer to modify the actor's behavior.
- (c) The number of *input* and *output ports* are assigned to an actor when it is created. The ports themselves are the location passed to the kernel routine when an input is to be received or an output is to be propagated. For example, a cross correlation actor may receive two blocks of time domain samples, one from each of the data streams to be correlated. Stream one is connected to one input port and stream two to the other. The result is propagated out the output port.

Thus, this actor has two input ports and one output port. An actor with no input ports is always ready for execution.

(d) The *control principle* determines what criteria will be used by the kernel scheduler to decide when an actor's script should be executed. There are currently two choices for this, *ifall* and *ifany*. Under *ifall* criteria, the actor will only be scheduled for execution after all input ports have data available. The *ifany* criteria causes the actor's script to be scheduled any time one or more data items are available for input.

2. *Real-time actornodes* are a special variation of actornode. They have two scripts. The first script is executed under the same criteria as a normal actor. Before completing, this first script should make a special kernel call to inform the kernel that the actor is waiting for a particular event to occur (such as an I/O channel communication to complete). The actor will be executed only after the event specified has occurred and then the second script will be used rather than the first.
3. *Datanodes* provide a queuing and connection function between actornodes. In order to allow certain portions of the graph to have a higher priority than others, the large-grain dataflow system implemented by the Multigraph kernel is asynchronous. The asynchronous nature of the graph forces the kernel to provide dynamic scheduling functions. As well, a way to queue input tokens between actornodes must be provided. This queuing function is provided by passive kernel components called datanodes. Any number of actor output ports may be connected to a datanode. A datanode may be connected to any number of actor input ports. Datanodes may be enabled or disabled. Unless enabled, a datanode will not allow data to be propagated through it to an actor's input port. Datanodes have been purposely left as a separate component from actors rather than providing the queuing function by way of an actor's input ports. Datanodes provide a simple interface for the symbolic layer to control, build, and monitor an executing graph. Functions to create, destroy, connect to actor input and output ports, disconnect from actors, clear, read, write, enable, disable, check length, set length, and check connections of datanodes are all provided for the symbolic interface.

4. *Environments* are used to protect system resources and provide a priority mechanism for sections of the dataflow graph. All actors are assigned to an environment. Each environment is given a numeric priority. The kernel scheduler always prefers actors connected to environments of a higher priority. Environments of the same priority are serviced in a round-robin fashion. Only one actor per environment will be executed at any one time. Thus, if two actors need to share a global resource, placing them in the same environment ensures mutual exclusion. Functions to create, destroy, attach to tasks, and set priority of environments are provided for the symbolic interface.
5. *Tasks* provide a generic interface to the basic computational resources of the underlying machine. In a multitasking environment they are simply the different processing threads available to the Multigraph kernel. In a multiprocessor system they are the individual processors themselves. Environments are attached to tasks. The *task* and *environment* concepts allow the actual implementation of the underlying system to be hidden from the user. Thus, the exact same model-based graph building techniques may be used regardless of the underlying hardware architecture. Functions to create and destroy tasks are provided for the symbolic interface.

In summary, the Multigraph kernel provides an interface to build, modify, monitor, and control an asynchronous macro-dataflow graph. The graph may be partitioned by the concept of environments and tasks. The actual dataflow graph looks slightly different from that of a classic dataflow graph due to the introduction of the passive nodes providing queuing functions, datanodes. The computation nodes are called actornodes. They have input ports, output ports, a context, a script, and a status. The use of distinct actornodes and datanodes provides the capability for the symbolic system to build, control, and monitor the dataflow graph.

SIGNAL FLOW GRAPH MODELING

In order to assist the creation and management of a particular dataflow graph, various tools and languages ranging from basic textual input to a graphical editing environment have been developed [3], [4], [5].

The Multigraph tools HDL and Xgem are frequently used to model signal processing systems. HDL stands for Hierarchical Design Language. The purpose of the language is to represent signal processing systems in a non-algorithmic manner. Rather, the system is described in terms of its *structure*. As such, HDL is classified as a *declarative* language. Block diagrams are often used to represent the signal flow among various modules of a complex signal processing system. The HDL declarations allow a simple means of describing these blocks and their interconnections. As well, HDL adds an extension to this approach through the concept of hierarchical decomposition. That is, any one block of the modeled signal flow diagram may itself be represented by a set of blocks (a sub-signal flow diagram). These hierarchies may proceed to arbitrary levels of abstraction. Complex blocks, made from one or more other blocks, are called *compounds*. Non-decomposable blocks are called *primitives*. The actions to be carried out by primitives are implemented with Multigraph *actornodes*. Each primitive has local parameters that have been inherited/derived through the hierarchy. After design of the top level signal flow graph of the hierarchy is complete, an interpreter (the Model Builder) converts it first to the primitive components of the hierarchy and then maps the result into the parallel execution environment provided by the Multigraph kernel. Each of the nodes of the resulting dataflow graph has a local state derived from the parameters inherited through the hierarchy (the actor's *context*). Node neighbors are known only through the input and output connection arcs. The resulting execution ready program is inherently parallel as a result of the declared signal flow graph being parallel. As well, it has been automatically generated from the high level declarations of the signal processing system and may now be scheduled by the underlying environment. Thus, the computer engineering issues around synchronization have been removed.

Xgem is a generic iconic editor program. It runs under X windows. Xgem has been tailored in correspondence with the HDL language. Through this interface, the user may build the declarations of his desired signal flow block diagram in an entirely graphical manner. Figures 2 and 3 show screen dumps from the Xgem environment.

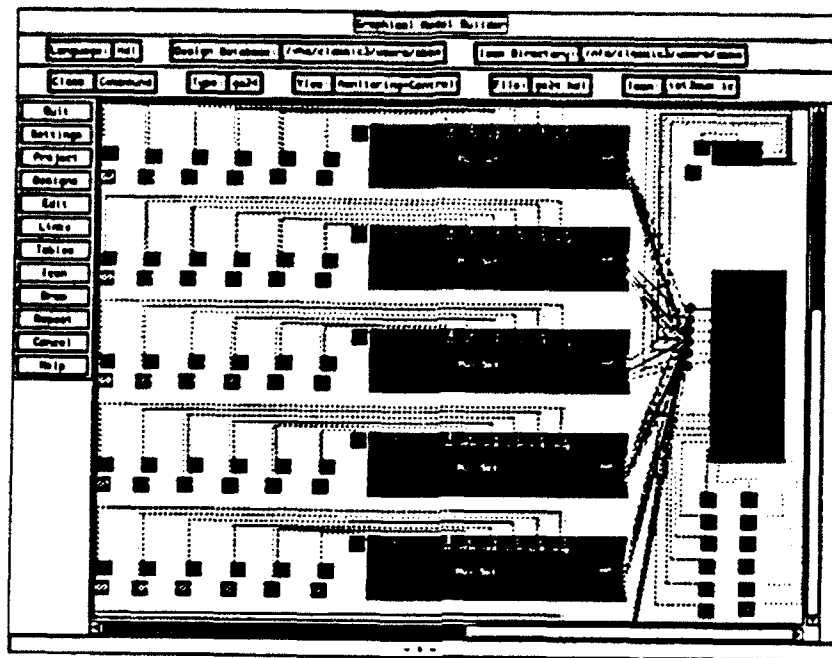


Figure 2: The top level CADDMAS signal flow graph

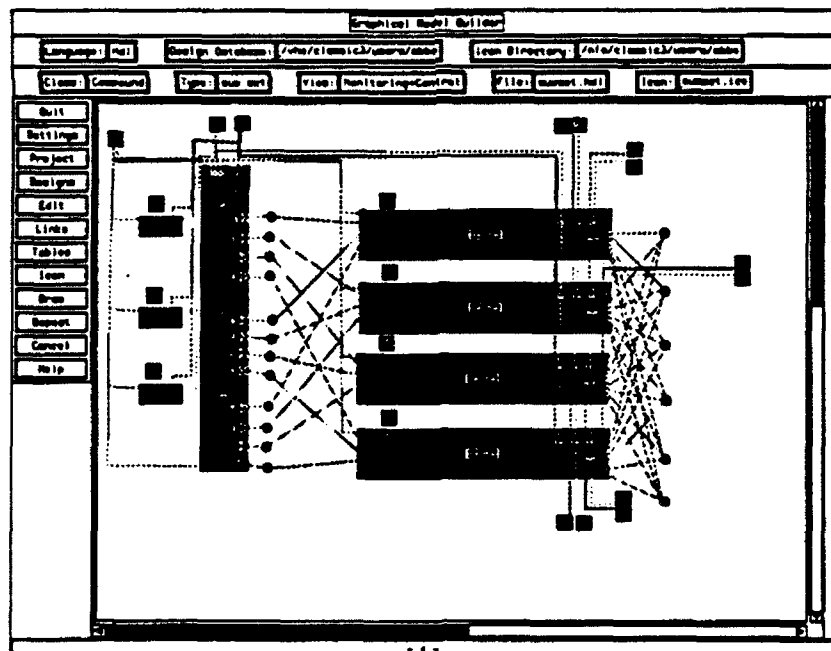


Figure 3: The CADDMAS multiplexor signal flow graph.

HARDWARE ARCHITECTURE MODELING

In order to assist the creation and management of a parallel architecture suitable for a particular application system, the Automatic Parallel Network Analyzer (APNA) tool has been developed. The basic computational blocks of parallel instrumentation systems often allow processor interconnections that are not fixed. For example, transputer architectures with DSP coprocessors fit signal processing applications especially well since the hardware connection architecture can be configured to match the structure of the signal flow graph.

As computational requirements increase, so does the need for more processors. When an architecture contains more than a couple of processors, managing hardware complexity becomes a serious problem. Testing to see if the network of processors has been interconnected correctly, searching for hardware errors, generating network information for an application loader, or creating map files for a message passing system by hand is time-consuming for tens of processors and impossible for hundreds.

Modeling the hardware architecture helps manage this complexity. Our approach again uses the generic graphical model builder Xgem. Models of parallel hardware architectures can easily be expressed in graphical terms (i.e. by icons and their interconnections). Hierarchy based modeling is used to manage complexity. The lowest hierarchy levels model the basic processors while the highest level constitutes the complete system. Intermediate levels model boards and subsystems. A number of attributes are attached to the processors (e.g. speed or memory size) in order to capture additional aspects of the system.

The declarative description of the hardware architecture is used to automatically produce load files and maps required by different software components of the system. The APNA tool loads the interpreted models and draws the network in a graphical window. It is capable of loading files of a number of different formats, comparing networks and displaying their differences, generating files for hardware diagnostic programs and a network loader, and displaying the attributes of the processors. It can selectively generate message passing maps, handling multiple paths between processors while optimizing the maps for minimum memory usage. Figures 4 and 5 show screen dumps from the Xgem used to model hardware architecture and the APNA tool.

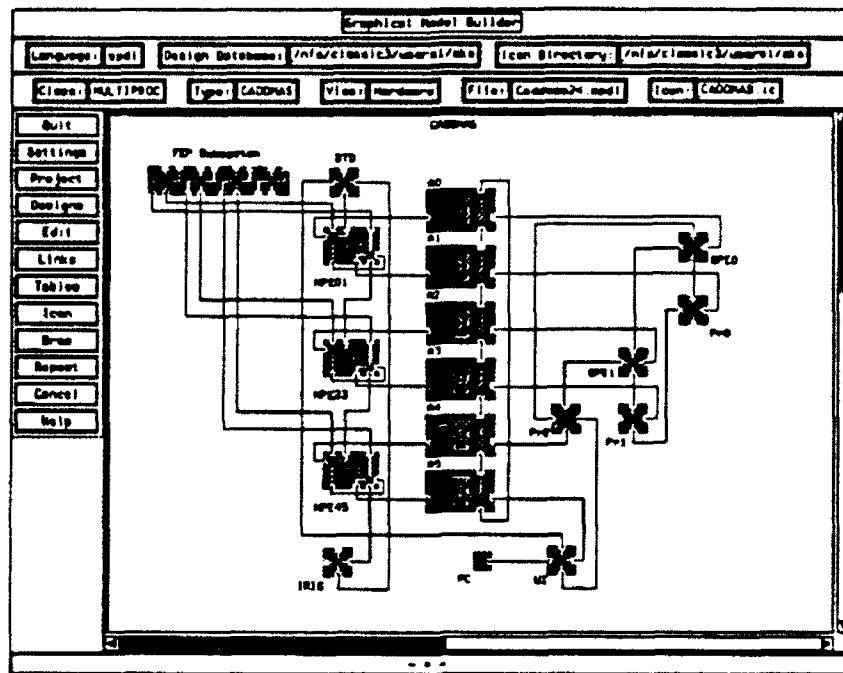


Figure 4: The top level CADDMAS hardware model set up in Xgem

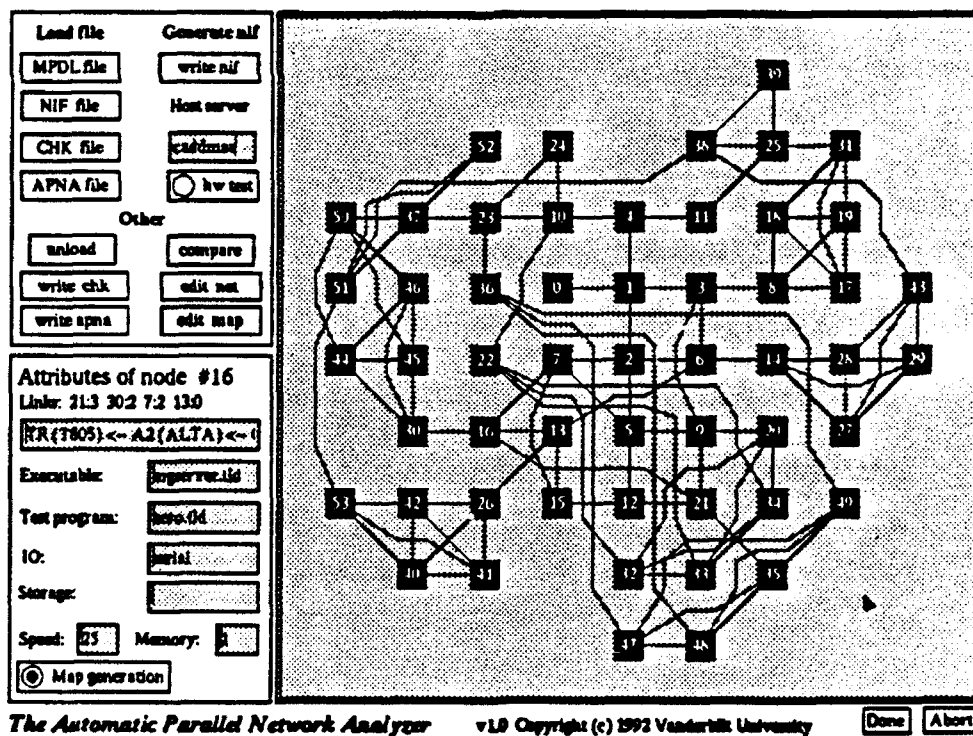


Figure 5: The APNA tool used to analyze the hardware model.

THE CASE STUDY PROBLEM: TURBINE ENGINE TESTING

Testing turbine engines involves running an instrumented version of the engine through various operational maneuvers (e.g. Acceleration or Throttle Snap). These tests are typically conducted while the engine is in a test cell (wind tunnel) capable of simulating altitude, atmospheric, and air speed conditions. In order to analyze dynamic vibrations, strain gauges (and other stress sensors) are attached to the turbine fan blades. A typical aeromechanic stress test instruments the engine with several hundred stress sensors along with a variety of temperature, pressure, flow, and revolution per second sensors. Stress sensors can generate signals with bandwidths in the tens of KiloHertz.

Historically, analysis of turbine engine stress data has been an off line process. On-line capabilities were limited to oscilloscopes showing unprocessed amplitude vs. time information and a small number of signal analyzers for simple spectral analysis on single channels. The bulk of the raw information was recorded onto analog tapes. Later, the analog tapes were digitized into conventional computers for analysis. The processing of this data was extremely compute intensive, and consequently, only a selected portion of the data was reduced. The analysis imposed a delay of several weeks on the availability of final results. Thus, vital information was not available for on-line test planning and evaluation.

The CADDMAS system was developed to provide these capabilities on-line. The system processes all sensor readings and presents the results both graphically and in hard copy form, during the test. The immediate availability of results opens the possibility for interactive test planning.

A graphical user interface allows the user to configure various visualization screens interactively. The user can select the number of visible windows on a screen, the contents of each plot window, and the parameters of each plot, such as titles, labels, axis ranges, and plot type, and display window update rate. Stored configurations automate the operation of the user interface. The user can also print any window or all windows on a screen.

COMPUTATIONS PERFORMED BY CADDMAS

The types of on-line plots generated by the CADDMAS system include:

- Amplitude vs. Time represents the absolute stress on a blade over time. Calculation involves scaling the sampled A/D counts to engineering units ($y = Ax + B$) and triggering the beginning of a display plot with a once per revolution pulse of the turbine.
- Spectral display with Full-Time Envelope Capture shows the energy of vibrations over a range of frequencies. The envelope continuously captures the historical maximum to ensure that no spectral activity is missed. FFT's are used to generate the spectral estimates.
- Frequency Tracking display shows the spectrum corrected by engine RPM. This is useful to observe synchronous vibrations over a range of engine speeds.
- Campbell diagrams are very important tools in the analysis of engine stress data. The Campbell diagram is a three dimensional scatter plot of stress amplitude vs. frequency vs. RPM. It represents historical information over an entire test maneuver and provides a summary of the blade behavior in a single chart. Generation of a Campbell diagram involves sorting the peaks of all of the spectral estimates of a strain gauge across a test maneuver and keeping the "interesting" peaks. Interesting is defined per test maneuver and includes: minimum, maximum, and delta RPM of interest, minimum, maximum, and delta frequency of interest, and minimum stress amplitude of interest.
- Phase Campbell diagrams show the phase angle relationship between two sensors vs. frequency vs. RPM. They incorporate information from two related sensors and can be used to help determine vibratory modes of an engine component.
- Modified Campbell diagrams allow the engineer to visualize the stress behavior of the engine components as a function of any engine or environmental parameter (instead of RPM) such as temperature, pressure, or time.

THE HARDWARE BUILDING BLOCKS

The CADDMAS hardware consists of various computational building blocks. INMOS transputers are used to provide a basic message passing fabric. Each transputer has four high-speed bidirectional serial communication links (20 Mega-bits per second) accessible through independent DMA engines. The current CADDMAS prototype system is capable of the on-line, gap free processing of 24 channels of stress, pressure, accelerometer, (and other) data at bandwidths up to 10 Kiloherzt. The bandwidth can be increased fourfold with a reduction in the number of channels multiplexed by the A/D's. The basic building blocks are modular so as to allow systems of various size and personality to be "plugged" together. These blocks include:

- The Front End Processor (FEP) consists of a high speed 12 bit A/D multiplexed to sample four independent analog streams, three Motorola 56001's are used for time domain and frequency domain threshold alarm generation, anti-alias filtering, and down sampling. The down sampled four data streams (to a 10 KHz bandwidth) are broken into 1024 point ensembles and shuffled through a FIFO into a parallel to INMOS serial converter chip, the C011, so they may be easily used as an input by other CADDMAS processing blocks.
- The Numeric Processing Element (NPE) consists of an INMOS T800 floating point transputer with a Zoran 34325 floating point Vector Signal Processor in shared memory. 38 MFLOPS are achieved when computing 1024 point FFT's on the Zoran. The actual packaging of the NPE is such that 2 independent NPE's share a single VME board.
- The Graphics Processing Element (GPE) consists of a Texas Instruments 34010 high-speed graphics processor with double buffered video RAM for smooth updates in shared memory with an INMOS T800 transputer.
- Basic T800 floating point processors are used as processing elements when the computations fit their general purpose architecture. Some of these blocks have parallel port adapters for driving printers or reading a special real-time clock.
- The User Interface (UI) and system builder capabilities are provided through Intel

based 486 PC's with INMOS serial to PC Bus adaptor cards.

Figures 2 through 5 show the CADDAMS signal flow graph and hardware architecture used for the current 24 channel prototype.

STEPS TO BUILD A NEW CADDMAS

To build a new CADDMAS system with a different personality, more channels, or different processing capabilities, the following procedure is used:

1. Model the desired signal flow graph using the HDL version of Xgem. For simply building a larger or smaller CADDMAS, the previously defined HDL compounds may be replicated without change. They simply need to be "plugged" together. If a new basic processing algorithm is needed, the script must be added to the Multigraph library and the corresponding HDL declarations added.
2. Use the Xgem tool to model the new, desired, hardware architecture.
3. Run the APNA system on the hardware model to produce: message passing maps, a network loader description file, and a hardware configuration diagnostic file.
4. Physically wire the new architecture. With a transputer system this involves plugging serial link patch cords into the appropriate sockets on the backplane.
5. Run the APNA configuration test program (Check) with its hardware configuration diagnostic file to find any wiring errors (there are about 100 patch wires in a 50 processor system).
6. Load and run the system. Starting the system causes the following actions to be performed:
 - (a) The APNA produced network loader description file is used to verify and load each of the transputers in the network with its necessary Multigraph kernel.
 - (b) The APNA produced message passing map is transferred down to each of the transputers in order to tell them what routes (including hops) to use to talk with one another.

- (c) The Xgem produced HDL models are interpreted by the model builder causing the individual nodes of the macro-dataflow graph, actornodes and datanodes, to be created and interconnected across their respective processors. In addition, the interpreter builds a database to be used by the interactive user interface allowing the signal flow graph to be manipulated by writing tokens into various datanodes of the graph.
- (d) The graph is now ready to be executed. Once started, the individual processors talk directly with each other as needed. The user interface can cause different graphs to be plotted at different rates and on different windows by writing various control parameters into particular datanodes of the graph.

EXPERIENCES AND FUTURE PLANS CONCERNING MODEL-BASED SYSTEMS

The model-based approach has proven to be a very useful tool to help manage the complexity of this large, parallel system. We have now built three different sized CADDMAS systems using the procedure outlined above. The first system was a four channel (single FEP), 8 Campbell, 2 graphics displays system. Actual configuration time for that first system could not be measured because we were still finishing some of the tools along the way. The second system was a 24 channel, 48 Campbell, 3 graphics displays, 3 laser printers system. It took one week to physically wire, and get the software running. That time included finding a previously missed bug in the message passing system and making several new patch cables. The third system was a four channel (single FEP), 4 Campbell, 1 graphics display system made from spare parts. It only took four hours to physically wire and get the software running! This proves that the approach is indeed flexible and efficient.

Future plans for the tools and techniques include: better methods to map the signal flow graph to the hardware, provide a dynamic load balance capability, automatically create and destroy pieces of the signal flow graph as required by user interface changes (this addition will greatly help the need for cross correlation operations since pre-building all possible cross correlations would require $(N * (N - 1))$ connections).

Future plans for CADDMAS applications include a 200 channel 50 KHz system using Texas Instruments TI 320C40 processors in conjunction with some of the current transputer

technology. The Multigraph, APNA, and HDL tools are sufficiently portable that the added processor type should not be a major problem.

ACKNOWLEDGEMENTS

The authors acknowledge the kind support of the United States Air Force in making possible the work described in this paper as well as the CADDMAS Team in addition to the authors:

- Capt. Greg Nordstrom: Arnold Engineering Development Center, Directorate of Technology.
- Tom Tibbals, Terry Hayes: Sverdrup Technology, AEDC Division
- Dr. Bruce Bomar: The University of Tennessee Space Institute

References

- [1] Sztipanovits, J., and Purves, R., "Coupling Symbolic and Numeric Computations in Distributed Environment", *Proc. of the Workshop on Coupling Symbolic and Numeric Computations*, Seattle, WA. pp.13/0-13/13, 1987.
- [2] Biegl, C.: "Design and Implementation of an Execution Environment for Knowledge-Based Systems" Ph.D. Thesis Dept. of Electrical Engineering, Vanderbilt University, Nashville, TN., Dec. 1988.
- [3] Karsai, G.: "Hierarchical Description Language (HDL) User's Manual" Dept. of Electrical Engineering, Vanderbilt University, Technical Report #87-004, 1987.
- [4] Karsai, G.: "Declarative Programming Techniques for Engineering Problems" Ph.D. thesis Dept. of Electrical Engineering, Vanderbilt University, Nashville, TN., Aug. 1988.
- [5] "Research on Intelligent Process Control Systems" Dept. of Electrical Engineering, Vanderbilt University, Technical Report #88-003, 1988.

SEE PREVIOUS REPORT FOR
COMPLETE TEXT.

**EXPERIENCES USING MODEL-BASED TECHNIQUES
FOR THE DEVELOPMENT OF A
LARGE PARALLEL INSTRUMENTATION SYSTEM**

**Ben Abbott
Csaba Biegl
Research Faculty
Theodore A. Bapty
PhD. Candidate
Department of Electrical Engineering**

**Vanderbilt University
400 24th Ave. S.
Nashville, TN 37235-1824**

**Final Report for:
Summer Research Program
Arnold Engineering Development Center**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

**DATA REDUCTION OF LASER INDUCED FLUORESCENCE
IN ROCKET MOTOR EXHAUSTS**

**Stephen H. Cobb
Assistant Professor
Department of Physics**

**Murray State University
Murray, Kentucky 42071**

**Final Report for:
Summer Research Program
Arnold Engineering Development Center**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

DATA REDUCTION OF LASER INDUCED FLUORESCENCE
IN ROCKET MOTOR EXHAUSTS

Stephen H. Cobb
Assistant Professor
Department of Physics and Astronomy
Murray State University

ABSTRACT

Laser induced sodium fluorescence data from a liquid-fueled rocket engine's exhaust plume have been reduced and analyzed. By correlating the experimental fluorescence linewidths with those of the theoretically predicted Voigt profiles, a temperature has been calculated as a function of radial position in the flow field. Also, by monitoring the Doppler shifted centroid of the fluorescence lineshape, the radial component of flow velocity has been determined. These experimental results are compared with computer code predictions.

DATA REDUCTION OF LASER INDUCED FLUORESCENCE
IN ROCKET MOTOR EXHAUSTS

Stephen H. Cobb

INTRODUCTION

The Arnold Engineering Development Center (AEDC) has developed a Laser Induced Fluorescence (LIF) system for non-intrusive diagnostics of high velocity exhaust flows. This system includes a tunable dye laser directed perpendicular to a flow field which contains atoms or molecules with known absorption properties over the frequency range of the laser. As the laser is tuned through the resonant absorption frequency of the atom or molecule, an accompanying fluorescence emission occurs which is recorded as a function of the laser frequency. The frequency dependence of this emission can be used to determine flowfield parameters such as temperature and velocity(1), and comparisons can be made to computer code predictions.

An experimental evaluation of LIF emission from a liquid fueled rocket engine was recently conducted by AEDC(2). The emitting atom in these studies was sodium, which occurs naturally in some high temperature exhaust systems(3). The fluorescence due to excitation near the D1 line (589.6 nm) was detected with a CCD camera and recorded on VHS tape. Each video frame was subsequently digitized and stored for reduction and analysis. This paper describes an initial approach which has been implemented in an effort to extract temperature and velocity

information from the LIF data. Also presented are preliminary results for this method of data reduction.

METHODOLOGY

Data Preparation

Details of the nature and geometry of the experimental apparatus and the method of data acquisition are given in reference 2, and will be mentioned here only briefly. During the videotaped motor firing each frame of video data, representing 33 msec real time, was digitized and stored on optical disk. A representative portion of this fluorescence data was selected as a data set by visually observing the onset and termination of fluorescence emission. This data set consisted of 83 frames of digitized information.

An integrated fluorescence intensity value was obtained by summing the response from all detector pixels located in a single column. This resulted in an intensity value for 512 pixel positions, corresponding to a particular radial position in the exhaust flow. These 512 values were recorded for each of the 83 frames of data. A typical plot of pixel intensity vs. radial position is shown in Figure 1, where the excitation laser frequency is near the D1 line frequency. All data points underwent correction for a baseline zero offset. Intensities were also adjusted to correct for the variable transmittance which occurred for light coming in at different angles of incidence on the sodium filter that fronted the CCD camera. The data were then stored in an array of 83 rows and 512 columns, the 83 rows representing the number of digitized frames of data, the

512 columns corresponding to the number of available detector pixels.

A second set of profiles were generated from this array. These profiles were created by monitoring the fluorescence intensity at a particular pixel, or equivalently, at a particular point in the flow, over the full 83 frame scan of laser frequency. These fluorescence profiles were smoothed over 5 points with constant weighting to yield a more familiar fluorescence lineshape (Figure 2). Subsequent analysis was performed on these profiles.

Data Analysis

The fluorescing specie in the exhaust flow field is, in general, subject to a variety of fluorescence line broadening mechanisms. The primary inhomogeneous broadening effect is Doppler broadening, where the atom's resonance absorption frequency is shifted depending on its velocity component in the direction of the probe laser beam. The lineshape associated with Doppler broadening is a Gaussian, whose full width at half maximum (FWHM) is proportional to the square root of the temperature. Line broadening associated with collisions is a homogeneous effect. The lineshape associated with collision (pressure) broadening is a Lorentzian, which tends toward zero more gradually than does the Gaussian.

In many cases, both Doppler and collision broadening must be taken into account in determining the linewidth. The lineshape resulting from a combination of the previously mentioned effects is called a Voigt lineshape. It can be expressed as (4):

$$V(a,x) = \frac{a}{\pi^{3/2}} \frac{1}{\gamma_D} \int_{-\infty}^{\infty} \frac{e^{-y^2} dy}{a^2 + (x-y)^2}$$

where $x = \nu - \nu_0$, and represents the detuning away from resonance, and $a = \gamma_L / \gamma_D$, the ratio of the Lorentzian width to the Doppler width.

The exhaust flow monitored in this experiment was known to be at a pressure of approximately 1 Torr. It follows that in this low pressure environment the parameter a will be very small. Thus the lineshape is expected to be dominated by Doppler broadening, and an experimental measure of the emission linewidth should infer a local temperature for the emitting atom. A detailed analysis of the conditions encountered in the flow was conducted by AEDC scientists (5). This analysis considered homogeneous broadening effects due to elastic collisions, quenching collisions, and spontaneous radiative decay and gave an estimate for the ratio of Lorentzian to Doppler widths of 1:62. This ratio was taken as the parameter a and was used to generate a Voigt lineshape centered on the $3P_{1/2} - 3S_{1/2}(F=2)$ transition in the atomic sodium system at a particular temperature. A second Voigt was generated having its center at a frequency higher by 1.77 GHz, corresponding to the $3P_{1/2} - 3S_{1/2}(F=1)$ transition. The two profiles were weighted by their relative transition probabilities, 0.625 and 0.375, respectively, and summed to produce the lineshape function expected for the fluorescence emission (Figure 3).

This process was repeated for temperatures ranging from 350 to 2000K. At each temperature the FWHM of the emission lineshape was measured. These linewidth-temperature correlations were then applied to the 512 fluorescence profiles which were similar to the one shown in Figure 2. The results are shown in the temperature vs. pixel location plot of Figure 4. Flow centerline is located at approximately pixel number 250.

The radial component of velocity of the flow field was determined by finding the Doppler shift in resonance frequency for the transition. The absorption frequency shift is given by:

$$\Delta\nu = \frac{v}{c} \nu_0$$

where v is the atom's velocity component in the direction of the probe beam, c is the speed of light, and ν_0 is the resonance frequency for the stationary atom.

The location of the centroid frequency for each of the 512 fluorescence intensity vs. frequency profiles was determined by integrating over the profile. If we assume that the flow field has no radial velocity component at centerline, frequency shifts may be determined relative to that point. Flow velocities were calculated from equation 2 and are plotted in Figure 5.

RESULTS

As stated earlier, each frame of video data captures information over a span of 33 msec. This video speed is a function of the equipment used to image the sodium fluorescence. For a probe laser scanning 30 GHz in 6 seconds, this time

interval translates into a frequency interval of 165 MHz per frame of video data. This uncertainty effects the linewidth measurements. If each frequency measurement is known to within ± 82.5 MHz, the calculated temperature is good to within ± 150 K as a worst case estimate of calculational error. Figure 4 also shows the radial temperature profile as determined by computer simulation. The code used in this simulation was a standard VIPER/SPF flowfield code. Note the relatively good agreement with experimental temperature values in the region near centerline. However, the experimental data trends toward much higher temperatures near the flow boundary. The code's omission of the thin, high temperature, subsonic boundary layer which lies along the nozzle wall may contribute to this effect.

Figure 5 compares the computer predicted radial velocity component with the experimental values. The experimental values are observed to track closely with the code predicted values. The velocity curve also has an inherent uncertainty. For sodium, the velocity Doppler shift is 1.7 MHz per m/sec. For an uncertainty in frequency of ± 82.5 MHz/frame, this gives a velocity uncertainty of ± 159 ft/sec.

To our knowledge, this experiment represents the first time that flowfield parameters have been measured from LIF data taken on an actual rocket motor firing. The methods of data reduction and analysis developed here may now be implemented relatively easily on the remaining data sets acquired in motor firings with similar performance characteristics. Also, with some modification, the methods may prove helpful in extracting

information from flowfields having significantly different properties.

REFERENCES

1. R.B. Miles, "Resonant Doppler Velocimetry," Phys. Fluids 18(6), 751, June 1975.
2. C.W. Brasier, "Application of LIF to Rocket Motor Exhausts," presented to 1992 Conf. on Advanced Earth-to-Orbit Propulsion Technology, Huntsville, AL, May, 1992.
3. E.W. Schmidt, Hydrazine and its Derivatives, John Wiley & Sons, New York, 1984.
4. See, for example, R. Coerdts and H. Gronig, "Determination of Pressure and Temperature of Gases by Parameter Fitting the FFT of an Absorption Line," Appl. Opt. 28(15), 3021, 1 August 1989.
5. Private communication with T.L. Wallace, Sverdrup Technology, Inc., AEDC, Tullahoma, TN.

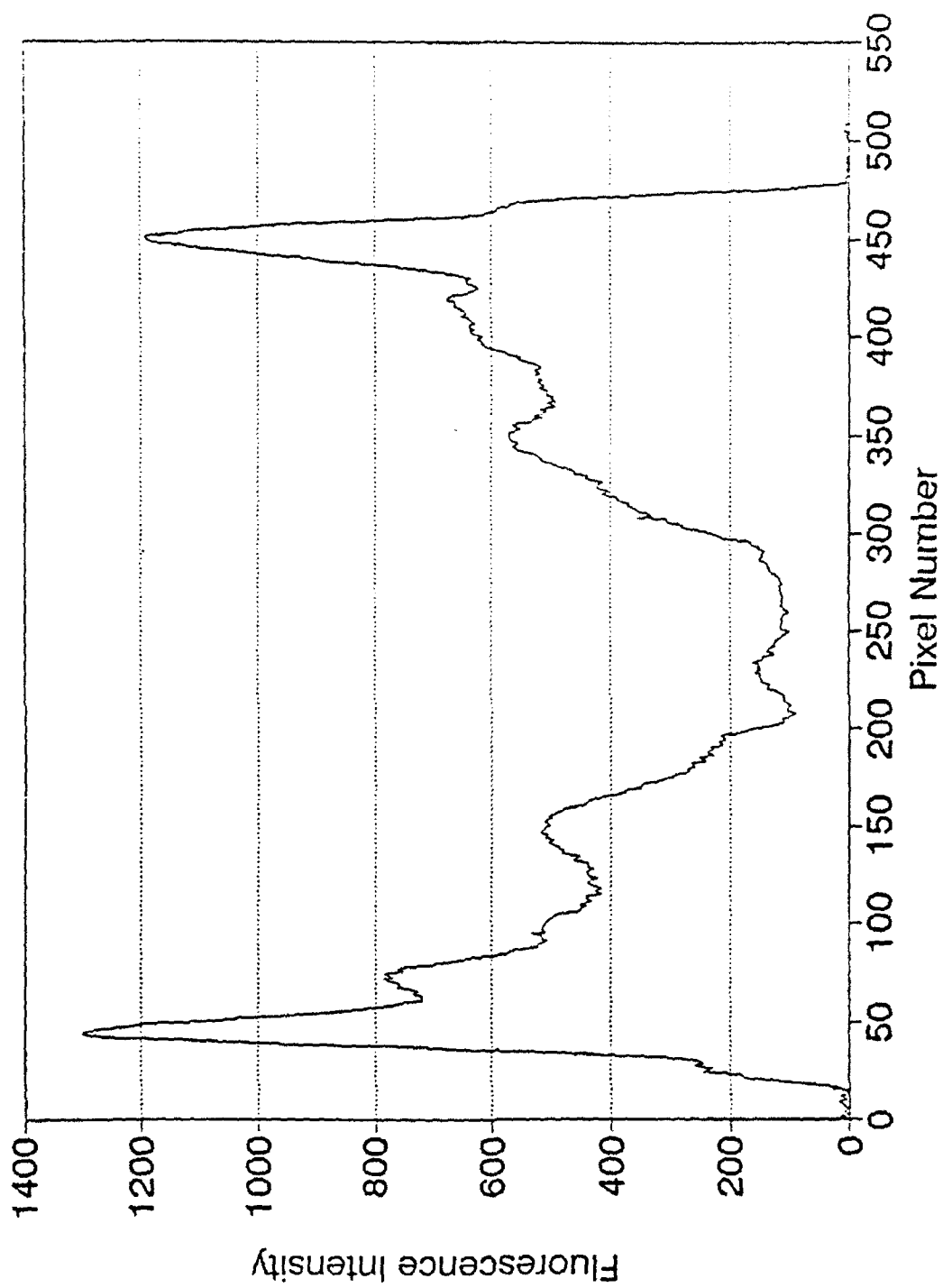


Figure 1. Fluorescent intensity of plume vs. radial position in the flow. Flow centerline is at approximately pixel number 250.

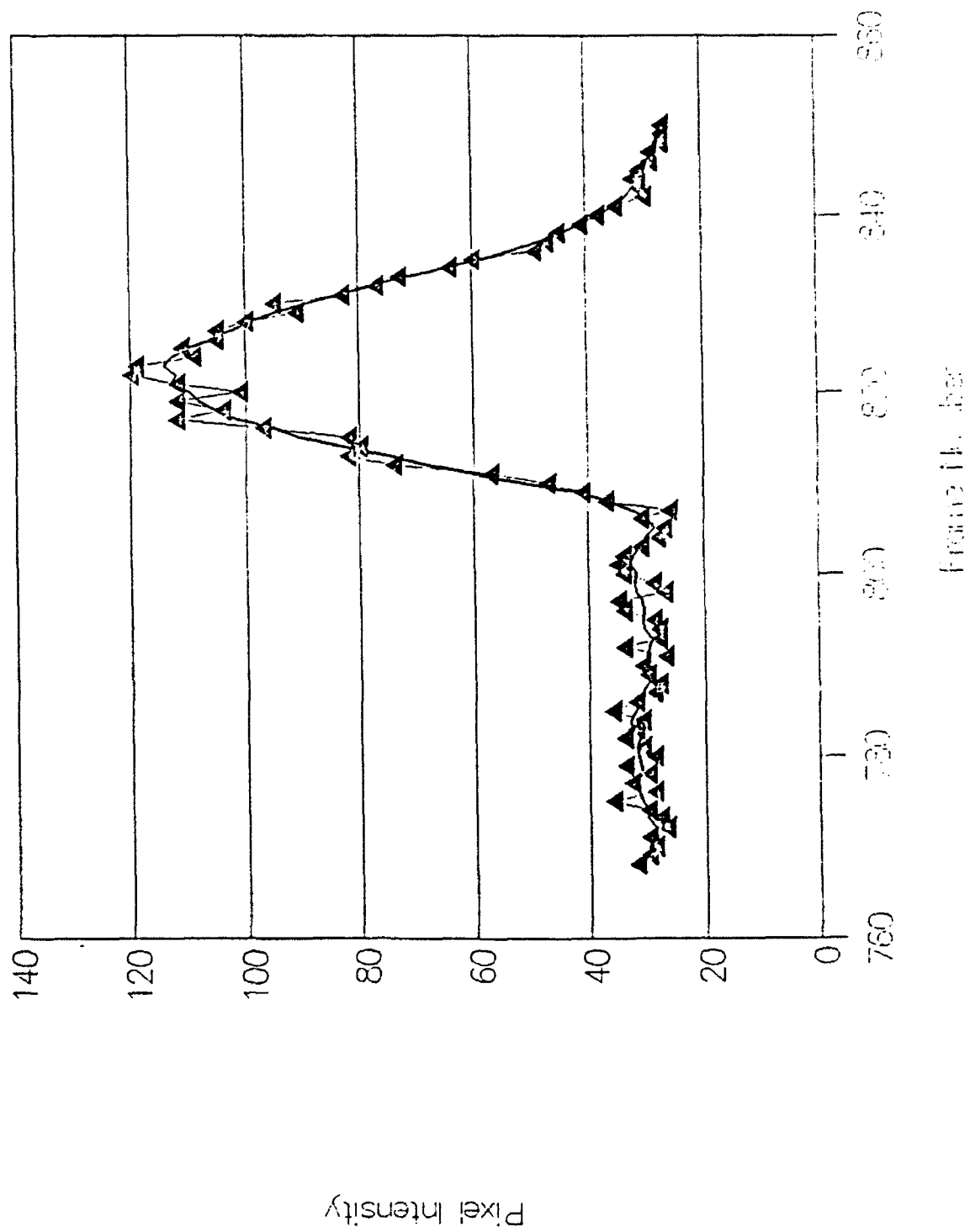


Figure 2. Example fluorescence profile of intensity vs. frame number, or frequency.

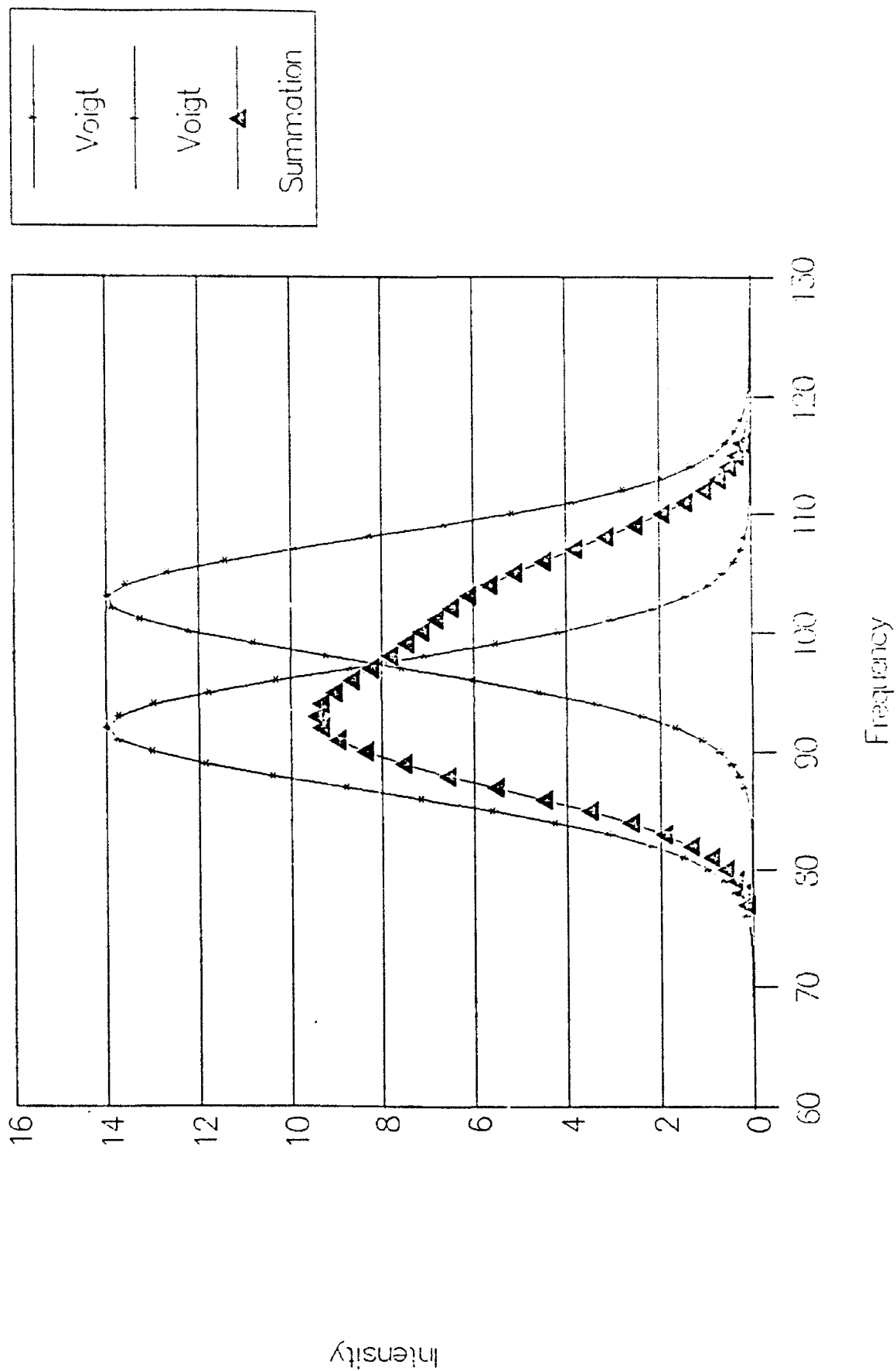


Figure 3. Two Voigt lineshapes separated by 1.77 GHz are weighted and summed to produce the fluorescence lineshape indicated by the triangles.

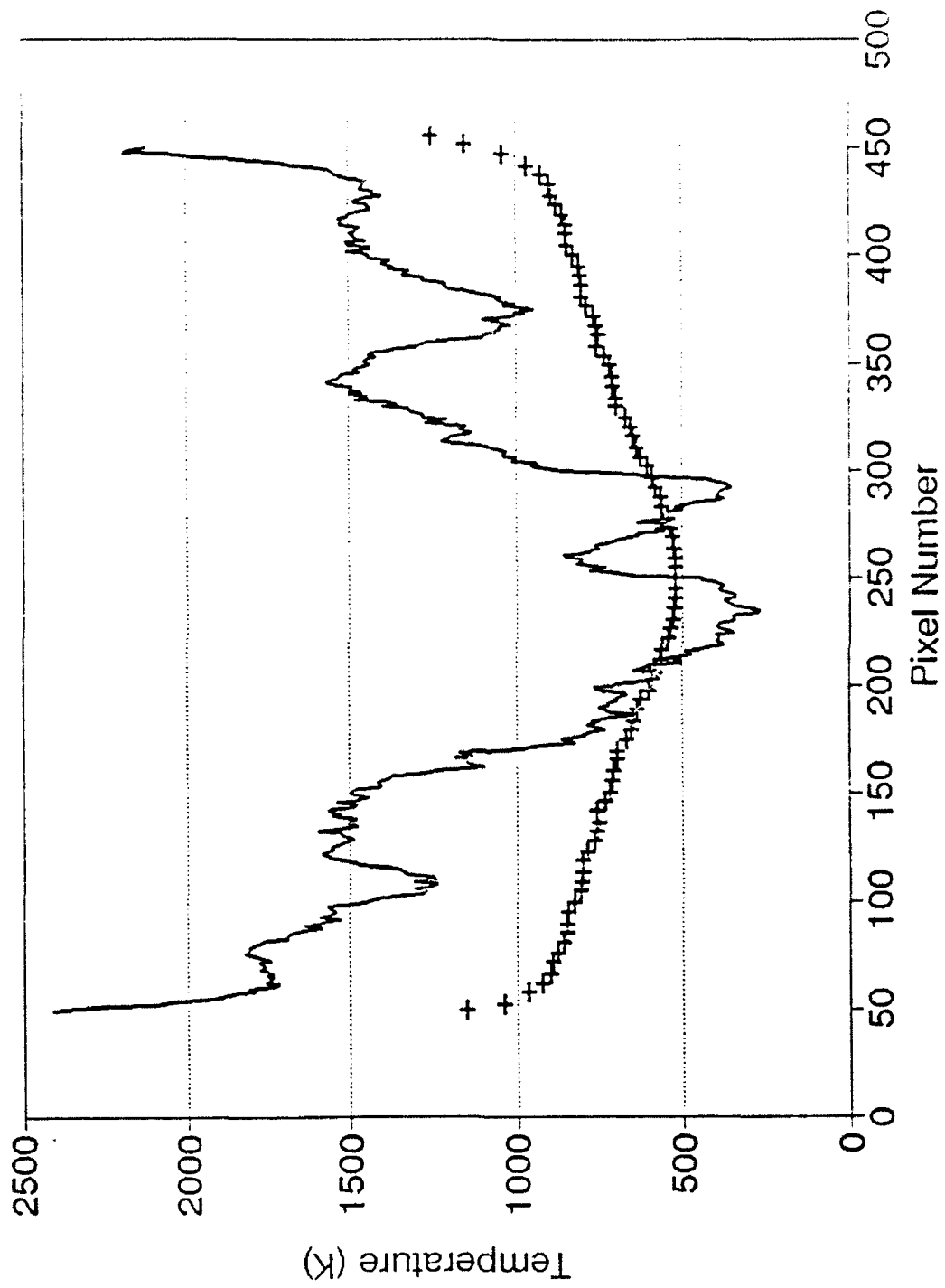


Figure 4. Temperature vs. radial position in the flow. Solid line is experimental data. Code predictions are indicated with + symbols.

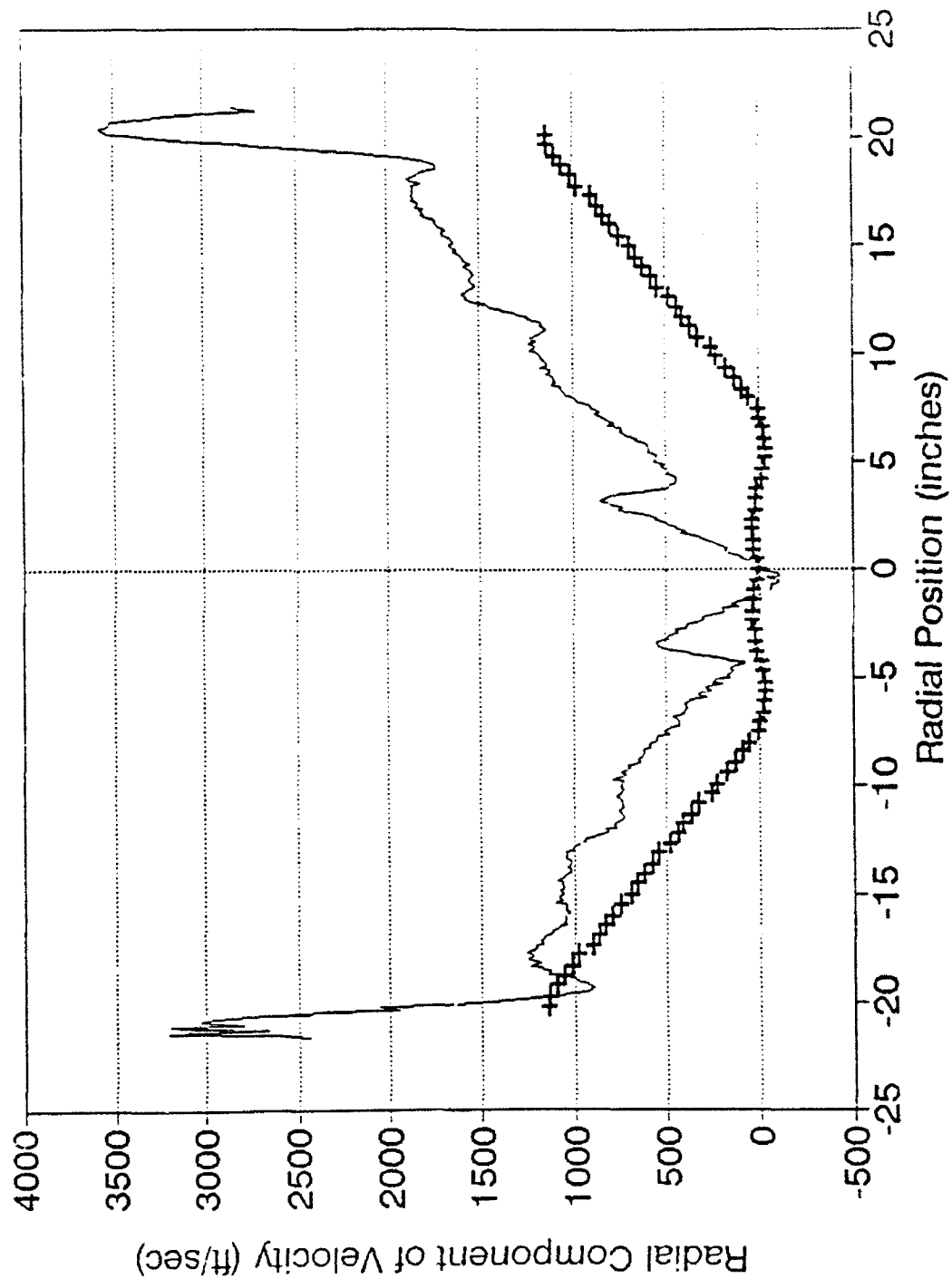


Figure 5. Radial component of velocity vs. position in the flow. Solid line is experimental data. Code predictions are indicated with + symbols.

**FEASIBILITY OF
WAVELET ANALYSIS
FOR PLUME DATA STUDY**

**Kenneth R. Kimble
Associate Professor
Department of Mathematics**

**University of Tennessee Space Institute
B. H. Goethert Parkway
Tullahoma, TN 37388-8897**

**Final Report for:
Summer Research Program
Arnold AFB**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

October 1992

**FEASIBILITY OF
WAVELET ANALYSIS
FOR PLUME DATA STUDY**

**Kenneth R. Kimble
Associate Professor
Department of Mathematics
University of Tennessee Space Institute**

Abstract

The recovery of information from images of rocket plumes presents severe challenges to image processing techniques. Severe noise is usually present since equipment is operated near the limit of its design. The images are non-stationary making Fourier analysis problematic. Classical non-Fourier analysis techniques require considerable manual judgement and intervention. Recent developments in an extension to Fourier analysis, "wavelet" analysis, and new techniques in fractal theory based on iterated function systems (IFS) both showed promise for developing automatic computer techniques which could treat both scale and dynamic variation in image data. Systematic investigation indicated that IFS is currently underdeveloped and the only published technique for image encoding utilizes an image grid segmentation which causes additional artifacts in the image at the grid lines. Wavelets, however, are applied without any such gridding, their encoding can be done with existing DSP hardware, and several classical image processing techniques can be implemented as adaptive wavelet bases. The analysis of textures in images has been found to be quite successful using wavelets and this analysis was attempted for rocket plumes. It was found that noise levels currently mask most texture in currently available images. However it was also found that wavelets form a more suitable tool than Fourier analysis for image understanding and that the potential exists to simplify other forms of image processing tools into wavelet filters.

FEASIBILITY OF WAVELET ANALYSIS FOR PLUME DATA STUDY

Kenneth R. Kimble

Introduction

The recovery of substantive information from digitized images of rocket plumes presents severe challenges to image processing techniques. The images typically contain large expanses of irrelevant detail in the form of star fields which lie around and behind the plume structure. The plume is recorded at low intensity levels with correspondingly low contrast levels within its structure, while it is this structure which is most important for the recognition of the characteristics of the flow. Often the plume structures occupy only a small percentage of the image because of the distance between camera and rocket. Cameras designed to capture images under adverse conditions produce relatively low resolution images with significant noise levels throughout the background. In the time domain the plume structure shifts rapidly within the image while star fields contribute erratic streaks to the background.

In the past few years two new analytic techniques have been developed which show promise in correcting, analyzing, and compressing digital images. One of these is based on fractal ideas²² and uses iterated function systems (IFS)³ to represent images in terms of adaptive fundamental images. The other, which can be viewed as a broad generalization of Fourier analysis, has been applied with success in signal processing¹² and image recognition¹ applications. This technique, called "wavelet" analysis^{5,6,7,9,21}, is specifically designed to treat non-stationary phenomena. In fact, there are indications of a deep relationship between these two techniques which is only hinted at by the state of the art in current research. Both techniques were considered during the past

summer.

Iterated function systems were developed originally by Barnsley². Very few substantive publications have been available until this past year. The IFS is promoted primarily as a technique for high ratio compression of images for application in telecommunications and high definition television. The IFS represents the image as a series of standard images (and associated functions) from a "code book" with associated weights. Images are reconstructed by applying the functions which correspond to the "code book" images to the points of the image until convergence is satisfactory. Constructing the codes for the images has become key to any successful application of the technique. The work published by Jacquin¹⁵ appears to be representative of methods for code construction; the image is first broken into small cells, then IFS codes are constructed on a best fit basis from a standard library of images. We believed that if the codes and their frequencies represented essential characteristics of the image, the codes could as well be used for image analysis as for the original compression purpose.

Fourier techniques have been used in image analysis for a long time. However unlike the Fourier technique whose spectral coefficients are all affected by any local time or space variation in the image, the wavelet component is only dependent on local time and space behavior. Localized techniques are not new, however, and most lose the significant advantage of high speed digital implementation which is a feature of the fast Fourier transform. But the FFT is paralleled in the case of wavelets by the existence of fast wavelet transforms (FWT) which are even slightly faster than the FFT. The FWT returns information in both scale (frequency) and time; however when the signal behavior is stationary, the time behavior is constant. Wavelets can be constructed with a variety of bases including ones specifically designed to match particular patterns in the signal; for example, in speech processing, individual phonemes may be matched by particular choices of basis functions.

Discussion of the Problem

Initially the problem was to determine whether either the iterated function technique or the wavelet techniques could be useful in analyzing rocket plume image data. A review of the literature and a critical analysis of the potential benefits, the level of effort in implementation, the availability of collateral research, and the general effectiveness of the techniques would need to be determined and compared. Then, in the time available in the summer, a pilot effort would be needed to illustrate a reasonable model of the application of the more promising technique to rocket plume analysis. There is current interest in attempting to characterize the texture which appears in plume images, and since texture analysis has both significant fractal dimensional characteristics as well as definite scale parameters which would be systematically represented by wavelet analysis, our emphasis was placed on the ability of the methods to detect and separate texture effects. In order to explain the direction of the project as it evolved, it is necessary here to explain the IFS in somewhat more detail.

The family of all images can be thought of as a (rather high dimensional) metric space (M, d) in which d represents the distortion metric between two images from M . The key to the IFS encoding of an image μ is to find a contractive mapping F which when applied iteratively to an arbitrary starting image from M will converge to the image μ . Barnsley has contributed significant research results to functional analysis in support of the underlying mathematical principles. One significant advantage of the IFS is that under broad conditions it is irrelevant whether one point (or a few points) is repeatedly transformed or all of the points of the succeeding images are transformed in parallel. This is an ideal algorithm for application of massively parallel as well as vector computer architectures.

The contractive mapping F is typically made up of a combination of affine transformations $f_1 \dots f_n$ with associated weights $w_1 \dots w_n$. The weights add to one and give the relative probability w_i that f_i will be applied at each step in the iteration. The f_i have so far been chosen largely on the

basis of experience; the ubiquitous fern is used in advertisements from a commercial company started by Barnsley since the fern is the "fixed point" of one the first f_i used to represent similar images of natural phenomena. Furthermore the ranges of the f_i are performed blockwise as explained in Jacquin. The image is broken up into non-overlapping cells and the range of the f_i is truncated to lie within the range block. Its domain, however, extends into neighboring blocks providing contextual sensitivity.

However, the IFS currently suffers some very serious disadvantages. Both the paper and the thesis by Jacquin^{14,15} provide an excellent starting point for the implementation of IFS based image analysis schemes which do provide a way to construct the IFS using a "code book" of images which might be chosen to match the problem domain, in this case, plume images. But the use of block segmentation of the image introduces serious artifacts and is undesirable in the already weak and noisy plume images. There is little information available on work done without segmentation. There is no guarantee of worst case convergence speed nor even effective practical measures for goodness of fit. As a result, there is no "vanilla" method which will be effective in a predictable way and from which adaptively selected f_i can be developed with a constant improvement in the resulting image representation. The f_i are only weakly related to familiar parametrizable characteristics. In particular, it is difficult to characterize particular textures even to their fractal dimension, much less to partition texture regions from the image.

The concepts of wavelet analysis developed from efforts to overcome the inability of Fourier analysis to treat non-stationary problems. Windowed Fourier transforms (WFT) were initially tried; indeed, finite Fourier transforms or FFTs are always a form of windowed transform. But the window of a windowed Fourier transform has constant width at all frequencies making it too small for low frequencies and too large for high frequencies. The continuous wavelet transform (CWT) is given by

$$(W_{\psi}f)(b, a) = |a|^{-\frac{1}{2}} \int f(x) \psi\left(\frac{x-b}{a}\right) dx$$

where ψ is a function which has no net disturbance, i.e. its integral is zero on the real line and it has finite energy there as well; consequently, ψ acts as a small wave or "wavelet". The CWT has a zoom-in zoom-out window owing to the presence of the scaling factor a in ψ and this overcomes the fixed window size of the WFT. As for the Fourier transform, the CWT is discretized to form a discrete wavelet transform (FWT) which has a natural form leading to fast algorithms²³ of equal or greater speed than the FFT. Rather than using the discretization, we will use a more appealing and direct theory that has been developed by Mallat^{19,20} which leads quickly to the FWT and its reconstruction and smoothing algorithms. The FWT does not suffer from the effects of blocking seen in IFS transforms; in fact, there are many choices of discrete bases which approximate the continuous function as closely as desired on the entire spatial and scale range. In addition extensive investigations have already been made into using adaptive basis functions to supplement the already complete standard bases.

Since the wavelet transform is already so well developed and so much applicable material exists, while the IFS transform is still in its infancy and totally unproven, we decided to base the pilot study on the use of wavelets. In the following we develop the wavelet transform algorithms for images in a manner following Mallat's multi-resolution analysis (MRA). We simulate a textured plume by constructing an artificial image with known texture scale and apply the wavelet transforms to it for a simple Haar¹³ wavelet basis and for the recently discovered Daubechies⁸ wavelet basis.

Methodology

We will derive the FWT based on the multi-resolution technique developed by Mallat. The MRA is based on six assumptions which seem to be a natural consequence of what one would expect of a systematic approach to analyzing conceptually continuous real world signals with discrete digital tools. We will state the assumptions and derive the WFT first in one dimension, then use a straight forward tensor product to obtain a two-dimensional FWT.

We consider approximations at discrete levels of accuracy; an approximation at scale j consists of selecting the best fit from a vector space which consists of "scale j " functions. From this it follows that we should consider a sequence of approximation spaces V_j which approximate finer scales as j increases. Specifically

$$\dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots \quad (1)$$

$$\bigcup_j V_j \text{ is dense in } L^2(R) \quad (2)$$

$$\bigcap_j V_j = \{0\} \quad (3)$$

This last ensures that approximation becomes coarse enough to include an arbitrarily long quasi-DC signal.

We will only consider changing scale in "octaves", i.e. we will scale by 2. CWT theory assures us that more than that gives redundancy in the approximation process. Whatever approximation we might make at one scale should be performed in an exactly similar manner at any other scale. Also we should find that approximation at two different points within the same scale are the same

except for translation. This is summarized in

$$f(x) \in V_j \iff f(2x) \in V_{j+1} \quad (4)$$

$$f(x) \in V_0 \iff f(x - k) \in V_0 \quad (5)$$

Finally we expect to use one fundamental measurement tool, translating it within a fixed scale, and scaling it to move from scale to scale. This can be axiomatized as

$$\text{There is a } \varphi(x) \in V_0 \text{ such that } \{\varphi(x - k)\} \text{ is an orthonormal basis for } V_0. \quad (6)$$

It follows from (1-6) that if we define

$$\varphi_{j,k}(x) = 2^j \varphi(2^j x - k) \quad (7)$$

then $\{\sqrt{2^{-j}} \varphi_{j,k}\}$ is an orthonormal basis for V_j .

Once we have specified the fundamental rules of the multi-resolution analysis, it becomes simple to describe the wavelets. Their function, as noticed by Mallat, is to provide a tool to measure the "detail j ", i.e. to measure the difference in the signal approximation at level j and level $j + 1$.

Specifically we define the sequence of spaces W_j to be the orthogonal complements of the V_j within V_{j+1} , i.e.

$$V_{j+1} = W_j \oplus V_j \quad (8)$$

Then it will follow that there is a "mother" wavelet ψ from which we can generate a basis for each of the W_j by defining

$$\psi_{j,k}(x) = 2^j \psi(2^j x - k) \quad (9)$$

from which $\{\sqrt{2^{-j}} \psi_{j,k}\}$ is an orthonormal basis of W_j .

Now we derive the filter which implements a FWT from the scale $j+1$ signal (called the average A_{j+1}) to obtain the average A_j signal and detail D_j signal. First note that since $\{\sqrt{2^{-j}}\varphi_{j,k}\}$ is a basis for V_j

$$(A_j f)(x) = 2^{-j} \sum_n \langle f, \varphi_{j,n} \rangle \varphi_{j,n}(x) \quad (10)$$

Now $\varphi_{j,n}$ is in V_{j+1} so

$$\varphi_{j,n} = 2^{-(j+1)} \sum_k \langle \varphi_{j,n}, \varphi_{j+1,k} \rangle \varphi_{j+1,k}(x) \quad (11)$$

Letting the FWT coefficient be denoted $A_{j,n}$ we define

$$A_{j,n} = \langle f, \varphi_{j,n} \rangle = 2^{-(j+1)} \sum_k \langle \varphi_{j,n}, \varphi_{j+1,k} \rangle \langle f, \varphi_{j+1,k} \rangle \quad (12)$$

The filter coefficient h_{k-2n} is given by

$$h_{k-2n} = 2^{-(j+1)} \langle \varphi_{j,n}, \varphi_{j+1,k} \rangle = \langle \varphi_{-1,0}(\xi), \varphi_{0,0}(\xi - (k - 2n)) \rangle \quad (13)$$

Hence the FWT for the scale $j+1$ to the scale j is a filter given by

$$A_{j,n} = \sum_k h_{k-2n} A_{j+1,k} \quad (14)$$

Similarly we obtain the transform from scale $j+1$ to detail j because the detail D_j is in W_j we have

$$(D_j f)(x) = 2^{-j} \sum_n \langle f, \psi_{j,n} \rangle \psi_{j,n}(x) \quad (15)$$

And because $\psi_{j,n}$ is in V_{j+1} it follows that

$$\psi_{j,n} = 2^{-(j+1)} \sum_k \langle \psi_{j,n}, \varphi_{j+1,k} \rangle \varphi_{j+1,k}(x) \quad (16)$$

So the FWT coefficient of the detail is denoted $D_{j,n}$ and becomes

$$D_{j,n} = \langle f, \psi_{j,n} \rangle = 2^{-(j+1)} \sum_k \langle \psi_{j,n}, \varphi_{j+1,k} \rangle \langle f, \varphi_{j+1,k} \rangle \quad (17)$$

Similar to the scale filter, the detail filter is

$$g_{k-2n} = 2^{-(j+1)} \langle \psi_{j,n}, \varphi_{j+1,k} \rangle = \langle \psi_{-1,0}(\xi), \varphi_{0,0}(\xi - (k - 2n)) \rangle \quad (18)$$

And the FWT for the scale $j + 1$ to the detail j is also a filter given by

$$D_{j,n} = \sum_k g_{k-2n} A_{j+1,k} \quad (19)$$

Both filters are "convolve and decimate" filters which are well known to the signal processing community. In addition, we may choose

$$g_n = (-1)^{1-n} h_{1-n} \quad (20)$$

which makes the detail filter a mirror of the scale filter. These filters may be applied recursively leading to a so-called pyramid algorithm to analyze the signal completely into its detail at all levels.

To obtain the image transform define $\Phi(x, y) = \varphi(x)\varphi(y)$, and $V_j^2 = V_j \otimes V_j$ so that Φ and $\{V_j^2\}$ forms an MRA for $L^2(R^2)$. Define

$$\Phi_{j,n,m}(x, y) = 2^{2j} \Phi(2^j x - n, 2^j y - m) \quad (21)$$

to obtain an orthonormal basis $2^{-j} \Phi_{j,n,m}(x, y)$. Then the mother wavelets are

$$\Psi^1(x, y) = \varphi(x)\psi(y), \quad \Psi^2(x, y) = \psi(x)\varphi(y), \quad \Psi^3(x, y) = \psi(x)\psi(y). \quad (22)$$

The translates and dilates of the Ψ 's are orthonormal bases for W_j^2 . The two-dimensional filters are obtained straightforwardly from h_j and g_j .

Results

In order to test the wavelet transform under controlled conditions, an artificial image of a rocket plume was constructed. A checkerboard pattern was chosen because it offered a simple means to specify the texture scale size and results would be pronounced enough to overcome the difficulties of printing images for this report in sufficient detail and clarity to make the behavior of the FWT evident. The detail visible on a work station screen is considerably greater than can be rendered here.

A true plume image is provided in Figure 1 for comparison; at present it is more a symbol of purpose than of accomplishment. Although it seemed at first to have texture which could be characterized, studies so far indicate uncorrelated noise at the finest scale, i.e. instrument noise, accounts for the appearance of texture. It may be possible to design successful wavelet based noise filters but this was not attempted in the present study.

The Haar wavelets which are essentially square wavelets were used to perform the FWT of Figure 3. Since these are a very good match to the checkerboard pattern and are well aligned with it, we see a best case result. The first level detail corresponding to Ψ^3 of the plume image appears in the upper right hand corner with Ψ^1 to the left and Ψ^2 below. The image is successively transformed leaving behind the detail and carrying the average down the diagonal to the left. In the lower left corner appears the final averaged image at the coarsest scale. Note that nearly the entire signal strength is concentrated at the third level down which corresponds to the Haar wavelet becoming a 4 pixel wide square wavelet signal. Both shape and orientation of the basis matches the image detail. This is what can be expected when a good analysis is achieved using the proper basis function matched to the detail sought and the coefficients are properly optimized.

The recently discovered Daubechies wavelets were used in the WFT which gave the results illustrated in Figure 4. The Daubechies wavelets are continuous and therefore do not switch as rapidly to

match the abrupt on-off characteristics of the checkerboard signal. So the detail components begin to appear at a finer scale than the signal actually contains. The "off-diagonal" contributions are also much more significant. Note, however, that very little distortion appears in the scales coarser than the signal. Continuous or higher order smooth basis in image analysis would normally be appropriate when the signal was similarly behaved; a good strategy is to use low order bases such as the Haar to remove sharp edged detail and to apply smoother filters only to model expanses of slowly varying image morphology.

Conclusions

Two new analysis techniques were investigated for potential application to the analysis of rocket plume data. Owing to a sparsity of theoretical foundation and to a lack of proven results, the IFS transform was not implemented. Implementation of the wavelet transform, though quite new in concept, was nevertheless accomplished by using filters in much the same manner as other signal processing techniques; indeed the image processing done here could be implemented straightforwardly on a suitably large capacity DSP chip for possible real time operation.

The preliminary results of the pilot study indicate that when the bases have elements which match the features of the image, very effective isolation of the contribution of the features to the image is possible.

The method here used to display the results, while compact and commonly used, is less satisfactory; only by choosing an artificial image with very pronounced behavior was it possible to show the success of the WFT in print.

Much remains to be done. Development of wavelet based noise filtering, more effective visualization methods, identification of adaptive functions to add to the bases in order to match rocket plume features, and incorporation of the time dimension into the wavelet transform are at the beginning of the list of possibilities. But there is little doubt that wavelet transforms will accomplish the first purpose for which they were invented, namely, to allow analysis of non-stationary phenomena as easily and effectively as Fourier transforms analyze stationary phenomena.

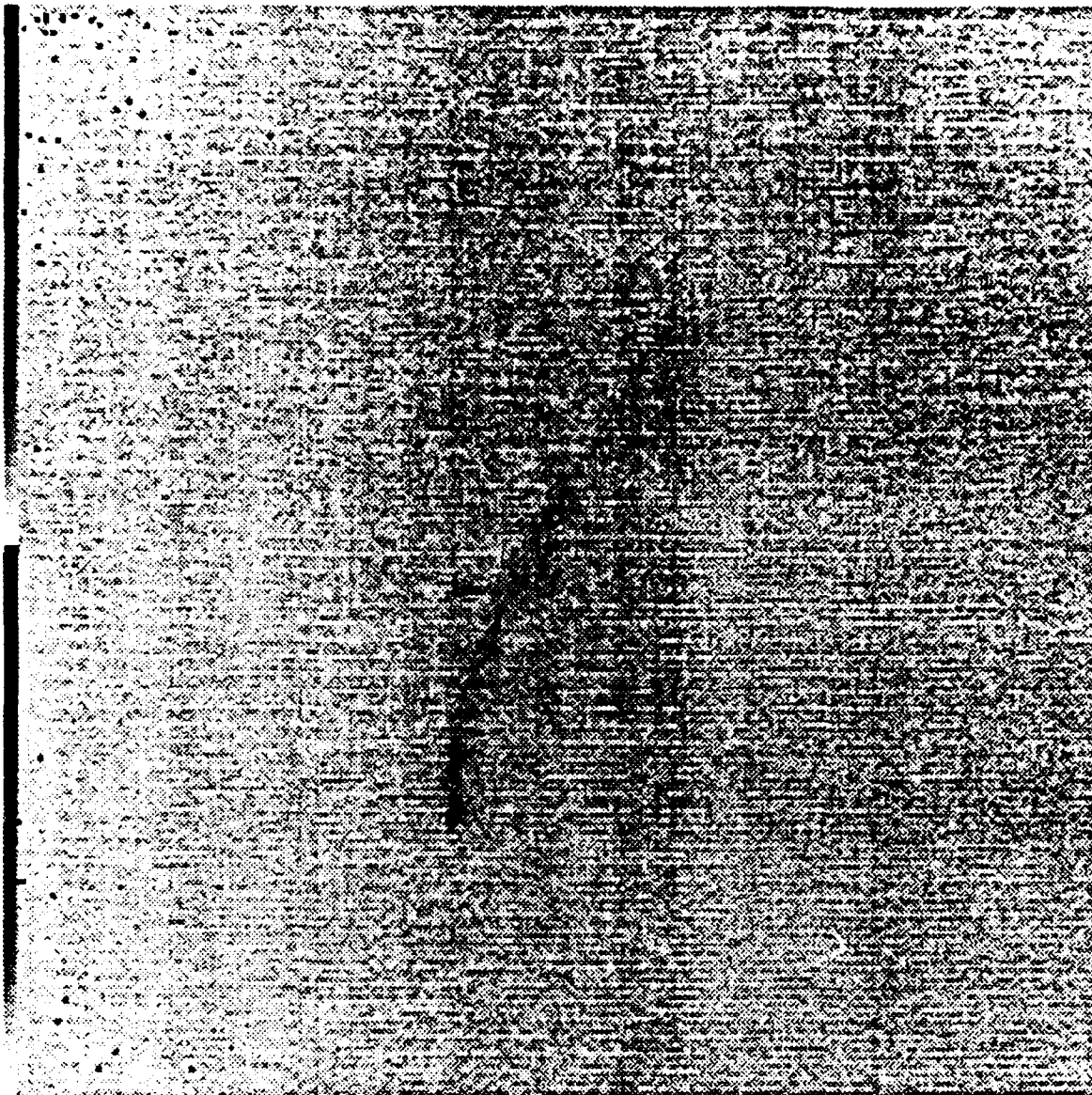


Figure 1. A typical 256 x 256 image of a rocket plume. Note the characteristic shape, high noise level, and weak image signal. (The image has been reversed to improve print quality.)

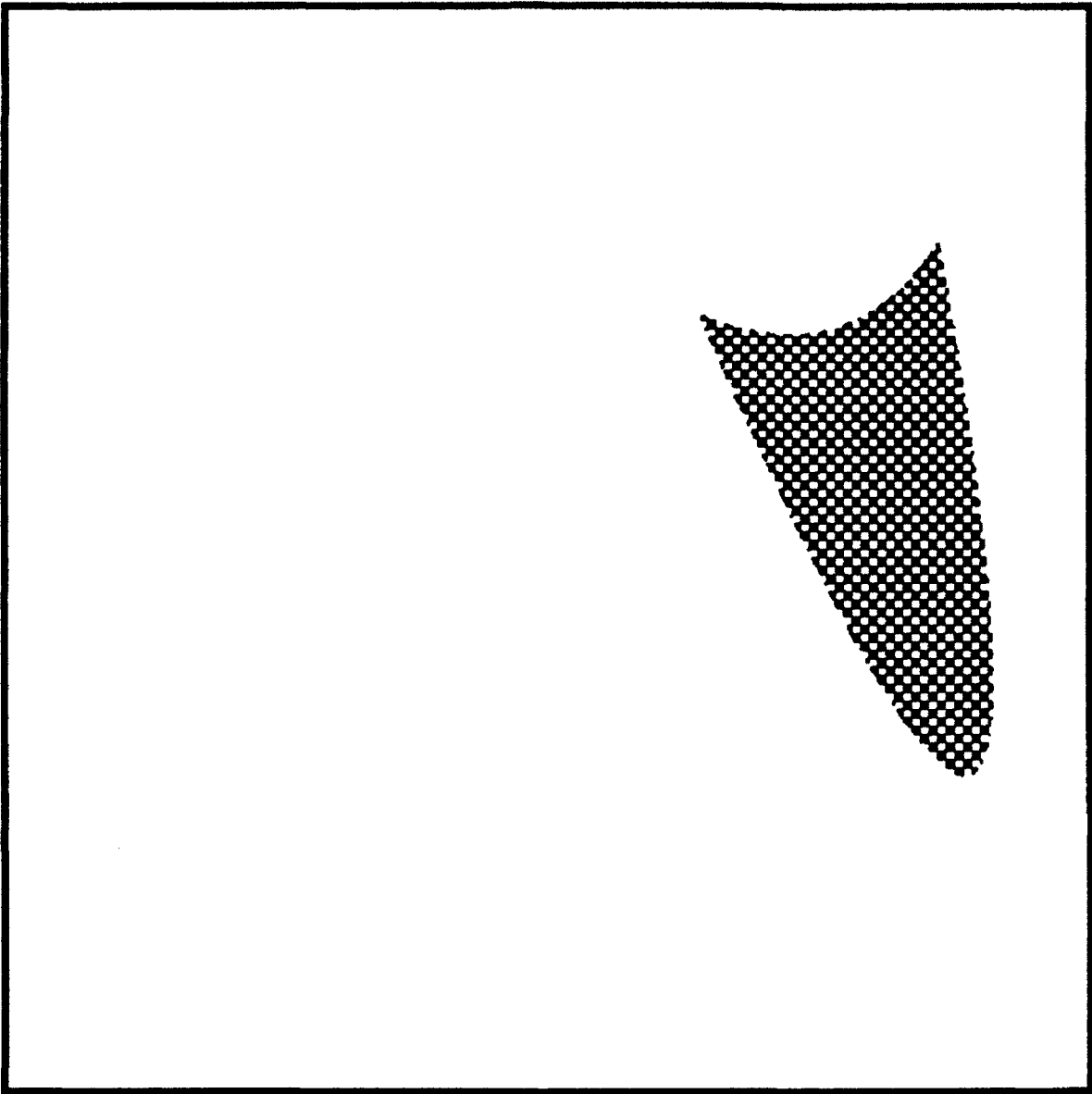


Figure 2. A simulation of a plume image with texture in the plume. The texture is created from a checkerboard of 4x4 pixels in a 512 x 512 image.

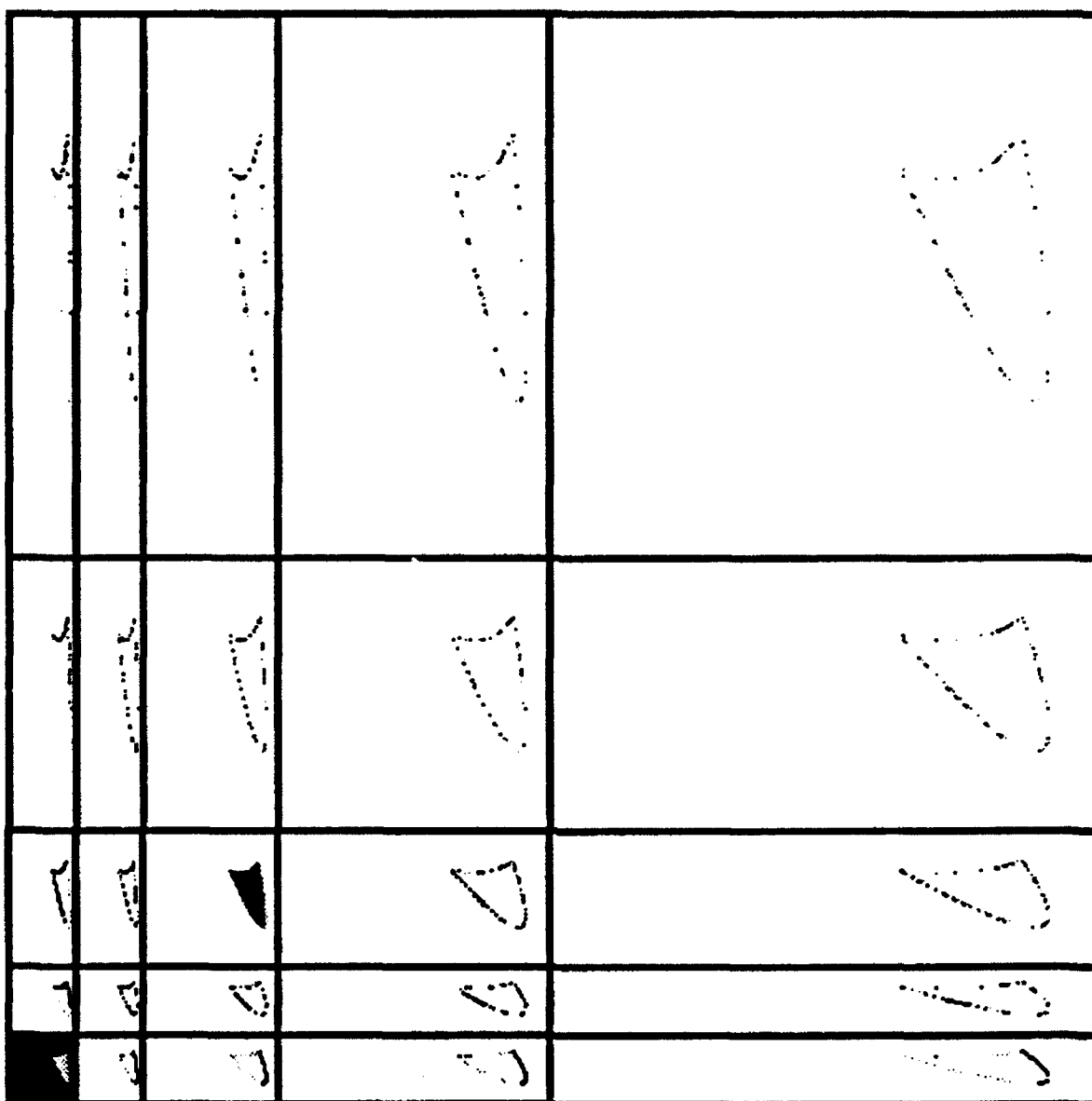


Figure 3. Wavelet analysis at several scales of the simulated plume using Haar wavelets. Note the nearly total absence of information in all scales except that corresponding to the texture scale.

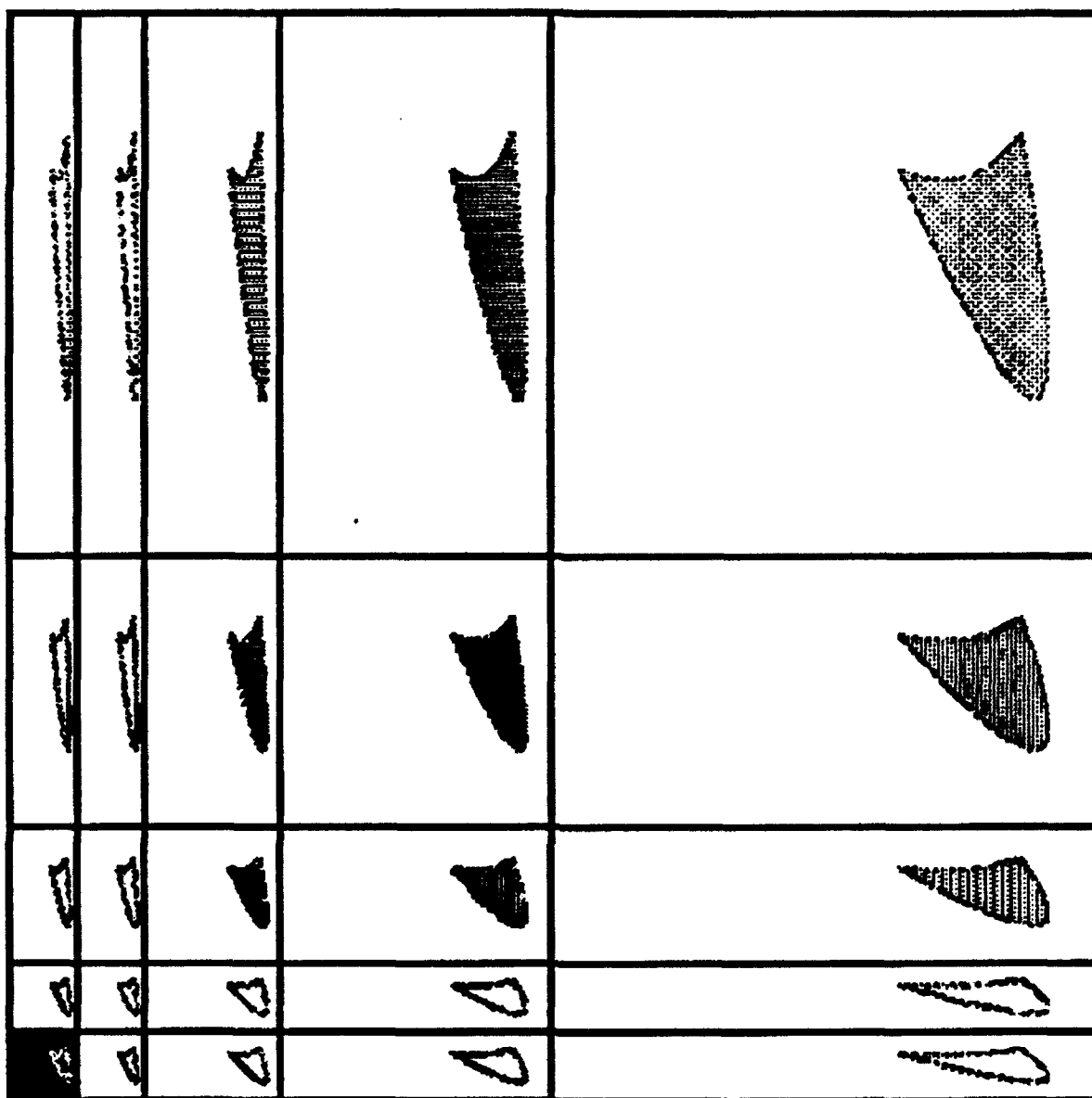


Figure 4. Wavelet analysis at several scales of the simulated plume using Daubechies wavelets.
 Note the gradation of information content as the scale of the the texture is approached.

References

- [1] J.P. Antoine, R. Murenzi, B. Piette, and M. Duval-Destin. *Image Analysis with 2D Continuous Wavelet Transform: Detection of Position, Orientation and Visual Contrast of Simple Objects*,. Proceedings of the International Conference on Wavelets, May 1989, Marseille, France; to be published by Masson, Paris, 1992.
- [2] Michael Barnsley. *Fractals Everywhere*. Academic Press, Inc., 1988.
- [3] Marc A. Berger. *Random Affine Iterated Function Systems: Curve Generation and Wavelets*. Society for Industrial and Applied Mathematics, 1992.
- [4] G. Beylkin, R. Coifmann, and V. Rokhlin. Fast wavelet transforms and numerical algorithms. *Comm. Pure Appl. Math.*, 44:141-183, 1991.
- [5] C. K. Chui. *An Introduction to Wavelets*,. Academic Press, New York, 1992.
- [6] C. K. Chui. *Wavelets: A Tutorial in Theory and Applications*,. Academic Press, New York, 1992.
- [7] Ronald R. Coifman, Yves Meyer, and Victor Wickerhauser. *Wavelet Analysis and Signal Processing*. Jones and Bartlett, Boston, 1992.
- [8] I. Daubechies. Orthonormal bases of compacity supported wavelets. *Comm. Pure Appl. Math.*, 41:909-996, 1988.
- [9] Ingrid Daubechies. *Ten Lectures on Wavelets*. Jones and Bartlett, Boston, 1992.
- [10] P. Dutilleux. An implementation of the 'algorithme à trous' to compute the wavelet transform. In J. M. Combes, A. Grossmann, and Ph. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase Space*, pages 298-304, 1989.
- [11] G. Fix and G. Strang. Fourier analysis of the finite element method in ritz-galerkin theory. *Stud. Appl. Math.*, 48:265-273, 1969.
- [12] P. Flandrin. Some aspects of non-stationary signal processing with emphasis on time-frequency and time-scale methods,. in *Wavelets*, J. M. Combes and A. Grossmann and Ph. Tchamitchian, eds., Springer-Verlag, Berlin, pages 68-98, 1989.
- [13] A. Haar. Zur theorie der orthogonalen funktionen-systeme,. *Math. Ann.*, 69, 1910.

- [14] Arnaud E. Jacquin. *A Fractal Theory of Iterated Markov Operators with Applications to Digital Image Coding*. UMI Dissertation Services, 1989.
- [15] Arnaud E. Jacquin. Image coding based on a fractal theory of iterated contractive image transformations. *IEEE Transactions on Image Processing*, 1, 1992.
- [16] J.L. Larsonneur and J. Morlet. *Wavelets and Seismic Interpretation*, in Proceedings of the International Conference on Wavelets, December 14-18, 1987, Marseille, France; to be published by Springer-Verlag, 1987.
- [17] W. Lawton. Necessary and sufficient conditions for constructing orthonormal wavelet bases,. *J. Math. Phys.*, 32, 1990.
- [18] W. Lawton. Tight frames of compactly supported wavelets,. *J. Math. Phys.*, 31, 1990.
- [19] S. Mallat. Multiresolution approximation and wavelets,. *Trans. Amer. Math. Soc.*, 315, 1989.
- [20] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI*, 11, 1989.
- [21] Y. Meyer. *Wavelets and applications*, in Proceedings of the International Conference on Wavelets, May 1989, Marseille, France; to be published by Masson, Paris, 1992.
- [22] Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe. *Chaos and Fractals*. Springer-Verlag, 1992.
- [23] Olivier Rioul and Pierre Duhamel. Fast algorithms for discrete and continuous wavelet transforms. *IEEE Transactions on Information Theory*, 38:569-586, 1992.
- [24] M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael. *Wavelets and their Applications*. Jones and Bartlett, Boston, 1992.
- [25] M. V. Wickerhauser. *Picture compression by best-basis sub-band coding*, preprint, Yale University, 1990.

**CURRENT FLOW IN THE PLASMA
EROSION OPENING SWITCH**

by

**Carlyle E. Moore
Associate Professor
Department of Physics**

**Morehouse College
830 Westview Dr., S.W.
Atlanta, GA 30314**

**Final Report for:
Summer Research Program
Arnold Engineering Development Center**

**Sponsored by:
Air Force Office of Scientific Research
Arnold Air Force Base, Tullahoma, TN**

September 15, 1992

CURRENT FLOW IN THE PLASMA EROSION OPENING SWITCH

by

Carlyle E. Moore
Physics Department
Morehouse College
Atlanta, GA 30314

ABSTRACT

The Plasma Erosion Opening Switch (PEOS) is a fast-opening switch that has been used successfully in a variety of pulsed power generators. It uses inductive, rather than capacitive, energy storage in order to exploit the high energy density associated with an inductor. A great deal of progress, both experimental and theoretical, has been made within recent years in describing the operation of the switch, and a number of approaches have been proposed in an effort to elucidate the underlying principles of switch behavior. While these provide a great deal of insight into the Physics involved, a comprehensive theory is yet to be found. This Report undertakes a review of some of the more important theoretical developments, with special emphasis on the dynamics of current flow in the Plasma, and examines them in light of experimental observations.

I. INTRODUCTION

The Plasma Erosion Opening Switch (PEOS) is a fast-opening switch that has been used successfully in a variety of pulsed power generators¹⁻⁵. It may be used to compress the power pulse from a conventional generator or as the primary switch in an inductive storage system. The use of inductive energy storage techniques for pulsed power has a number of advantages over conventional, capacitive techniques, the most important being the very high energy densities which can be obtained. In this report, we shall be concerned exclusively with inductive energy storage.

The structure of a typical PEOS system is shown schematically in Fig. 1. Plasma is injected between the electrodes of a coaxial transmission line shortly before the onset of the generator current, which is delivered to the storage inductor through the generator impedance. The operation of any opening switch may be conveniently divided into two basic stages: conduction and opening. During conduction, the switch behaves essentially like a short circuit, isolating the load from the generator. At a current level which is determined by the characteristics of the system, the plasma ceases to conduct (the switch opens) and the current is rapidly diverted to the load. The overall effect of the PEOS is to modify the output of the

generator in such a way as to deliver a similar current, higher voltage and a shorter-duration pulse to the load than could be accomplished by the generator alone (with matched load). Experiments demonstrating the PEOS technique have been performed at the Naval Research Laboratory (NRL), using a variety of generators⁴⁻⁷. In the earliest of these experiments, performed on the GAMBLE I generator, the plasma switch conducts the generator current for 50 ns as it rises to 200 kA, then opens and transfers the current to the (short-circuit) load in less than 10 ns. With an 11 Ω diode load, 1.4 MV is generated, more than twice the matched load voltage. These results are shown in Fig. 2. Subsequent experiments in the 50 ns conduction-time range were performed on the GAMBLE II generator, producing much higher

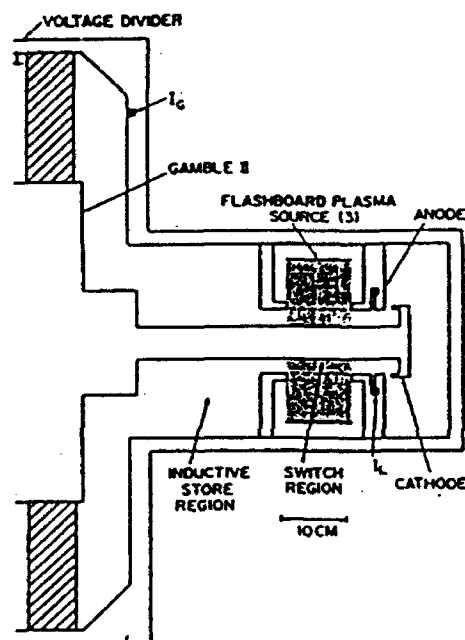


Fig. 1. Typical PEOS system showing inductive storage and plasma injection

currents and output voltages. Conduction times of $1 \mu\text{s}$ have been achieved with the HAWK generator, producing a current of 650 kA, a load voltage of 750 kV and 0.37 TW peak power⁵.

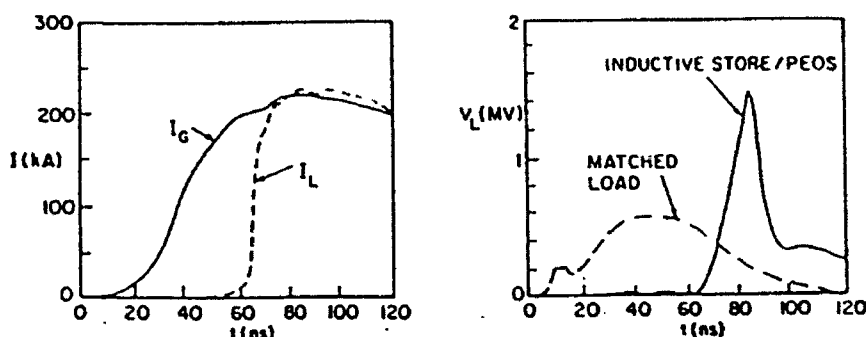


Fig. 2. GAMBLE I results. (a) Generator current I_G and load current I_L with short-circuit load. (b) Load voltage and matched voltage with diode load.

A number of approaches have been developed which provide insight into the operation of the PEOS. System requirements for pulse compression and power multiplication may be analyzed within the framework of a simple lumped circuit¹, in which it is assumed that the switch resistance is initially zero, but rises instantaneously at the end of the generator voltage pulse to a finite value, at which point the switch is opened and current flows to the load. This is discussed in Section II. The operation of the PEOS has been described¹⁻⁷ as a sequence of four distinct phases: conduction, erosion, enhanced erosion and magnetic insulation, the last three phases comprising the opening stage of the switch. The four-phase theory is discussed in Section III.

A thorough understanding of the operation of the PEOS cannot be gained, however, without detailed study of the dynamics of current flow in the plasma, including the Physics of space-charge sheaths and the relativistic orbits of electrons in electric and magnetic fields. Theoretical analysis of the problem is rather difficult, and relies, in the main, on analytical models and numerical simulations of switch behavior. In Section IV, we describe an electromagnetic two-fluid code⁶, which identifies two mechanisms, viz. plasma compression and magnetic insulation of electrons at the cathode, which play an important role in the conduction process. In parallel with this undertaking, a theoretical analysis is carried out, in which the fields derived from the fluid code are used and the fluid equations for the plasma ions are supplemented by the bipolar space-charge limited flow condition and a number of other simplifying assumptions. The degree of plasma compression (due to $\mathbf{J} \times \mathbf{B}$ forces) is obtained from the evolution of the ion density profiles. Magnetic insulation is examined by

deriving the orbits of test electrons which are accelerated across the cathode sheath and injected into the plasma with energies corresponding to the plasma potential.

II. LUMPED CIRCUIT ANALYSIS

System requirements for pulse compression and power multiplication may be analyzed in terms of the simple lumped circuit¹ shown in Fig. 3. The generator delivers a current I_1 to

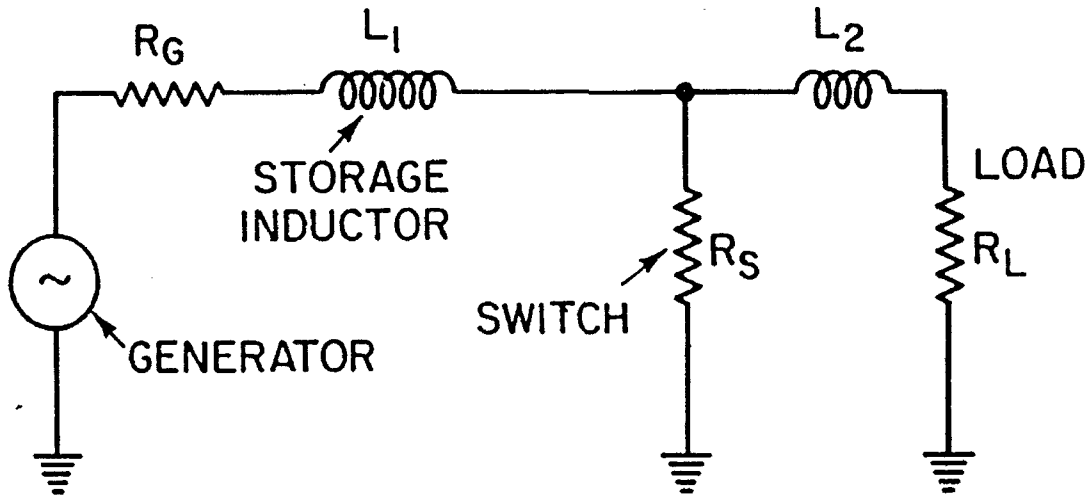


Fig. 3. Lumped circuit for analyzing system requirements of PEOS.

the storage inductor L_1 through the generator resistance R_G and the switch resistance, which is assumed to be zero (thus isolating the load). At the end of the voltage pulse produced by the generator (this is taken to be time $t = 0$), the storage inductor current is taken to be $I_1(0) = I_0$. At this point, the switch resistance rises instantaneously to R_S (switch is opened) and current flows to the load. The open circuit voltage across the generator is taken to be zero during this process. The load current by I_2 flows through the load resistance R_L and the parasitic inductance L_2 between the switch and the load. The circuit equations are

$$-L_1 \dot{I}_1 = R_G I_1 + R_S (I_1 - I_2) \quad (2.1)$$

$$-L_2 \dot{I}_2 = R_L I_2 - R_S (I_1 - I_2) \quad (2.2)$$

The equations may be uncoupled by taking the time derivative of (2.1) and substituting for $\dot{I}_2$ from (2.2) and I_2 from (2.1). This gives

$$\ddot{I}_1 + \left[\frac{R_G + R_S}{L_1} + \frac{R_L + R_S}{L_2} \right] \dot{I}_1 + \left[\frac{(R_G + R_S)(R_L + R_S) - R_S^2}{L_1 L_2} \right] I_1 = 0 \quad (2.3)$$

whose solution is of the form $I_1 = e^{-\alpha t}$, with $\alpha^2 - b\alpha + c = 0$, where

$$b = \frac{R_G + R_S}{L_1} + \frac{R_L + R_S}{L_2} ; c = \frac{(R_G + R_S)(R_L + R_S) - R_S^2}{L_1 L_2}$$

$$\text{Thus } \alpha_{\pm} = \left[\frac{R_G + R_S}{2L_1} + \frac{R_L + R_S}{2L_2} \right] \left[1 \pm \left(1 - \frac{4L_1 L_2 \{ (R_G + R_S)(R_L + R_S) - R_S^2 \}}{\{ L_1(R_L + R_S) + L_2(R_G + R_S) \}^2} \right)^{1/2} \right] \quad (2.4)$$

We assume that the storage inductor L_1 is much greater than the parasitic inductor L_2 , and that the switch resistance R_S is much larger than R_L or R_G (which allows the current to switch).

We may therefore make the approximation

$$\alpha_{\pm} = \frac{R_S}{2L_2} \left[1 \pm \left\{ 1 - \frac{2L_2(R_G + R_L)}{L_1 R_S} \right\} \right] \quad (2.5)$$

or
$$\alpha_+ = \frac{R_S}{L_2} ; \alpha_- = \frac{R_G + R_L}{L_1} \quad (2.6)$$

and we see that $\alpha_+ \gg \alpha_-$. Note the symmetry of Equations (2.1) and (2.2). The latter equation may be obtained from the former by making the substitutions $L_1 \rightarrow L_2$, $I_1 \rightarrow I_2$ and $R_G \rightarrow R_L$. Clearly, the solution of (2.2) is also of the form $e^{-\alpha t}$, since α is invariant under the changes indicated above. We therefore write

$$I_1(t) = A_1 e^{-\alpha_+ t} + B_1 e^{-\alpha_- t} \quad (2.7)$$

$$I_2(t) = A_2 e^{-\alpha_+ t} + B_2 e^{-\alpha_- t} \quad (2.8)$$

The boundary conditions $I_1(0) = I_0$ and $I_2(0) = 0$ give

$$A_1 + B_1 = I_0 \quad (2.9)$$

$$A_2 + B_2 = 0 \quad (2.10)$$

Substituting (2.9) and (2.10) into (2.1) and (2.2) and comparing coefficients of $e^{-\alpha_+ t}$ and $e^{-\alpha_- t}$, we get

$$A_1 = A_2 = \left(\frac{L_1}{L_1 + L_2} \right) I_0 \quad (2.11)$$

$$B_1 = \left(\frac{L_2}{L_1 + L_2} \right) I_0 ; B_2 = - \left(\frac{L_1}{L_1 + L_2} \right) I_0 \quad (2.12)$$

where we have used (2.6) and the inequalities $L_1 \gg L_2$ and $R_S \gg R_G$. It will now be more convenient to use the notation $\alpha_{\pm} = 1 / \tau_{\pm}$, and write the final solutions in the form

$$I_1(t) = \left(\frac{L_1}{L_1 + L_2} \right) \left(e^{-t/\tau_-} + \frac{L_2}{L_1} e^{-t/\tau_+} \right) I_0 \quad (2.13)$$

$$I_2(t) = \left(\frac{L_1}{L_1 + L_2} \right) (e^{-t/\tau_-} - e^{-t/\tau_+}) I_0 \quad (2.14)$$

The switch current $I_s = I_1 - I_2$ is then given by

$$I_s(t) = I_0 e^{-t/\tau_+} \quad (2.15)$$

and the decay time of the switch current is just $\tau_+ = L_2 / R_s$. We see therefore that small L_2 and large R_s will lead to rapid switch opening. Since $\tau_+ \ll \tau_-$, the pulse width of the load current is approximately equal to τ_- , and the ratio of the input pulse width $\tau_p = L_1 / R_G$ to the output pulse width is $\tau_p / \tau_- = (R_G + R_L) / R_G, > 1$, which demonstrates the pulse compression produced by the switch. Using the expression for the load current obtained above, it can be shown¹ that power multiplication is also achieved (relative to the matched load power), with a power multiplication factor M given by

$$M = \left(\frac{4R_L}{R_G} \right) \left(\frac{L_1}{L_1 + L_2} \right)^2 \left[1 - \exp \left(-\frac{R_G \tau_p}{L_1} \right) \right]^2 \left[1 - \frac{2}{x} \{ \ln(x) + 1 \} \right] \quad (2.16)$$

where
$$x = \frac{R_s L_1}{L_2 (R_G + R_L)} \quad (2.17)$$

Thus high power multiplication is achieved if the load resistance R_L is much greater than the generator impedance R_G .

III. THE FOUR-PHASE THEORY

While the lumped circuit analysis given above demonstrates the potential of the switch to produce pulse compression and power multiplication, it bypasses the important details of switch conduction. The observed behavior of the PEOS has been described in terms of a theory¹⁻⁷ developed at NRL, in which the operation of the switch is envisaged as a sequence of four distinct phases, viz. conduction, erosion, enhanced erosion and magnetic insulation. The plasma is injected into the vacuum coaxial region between the anode and the cathode a few microseconds before the generator is fired, thereby bridging the gap between the electrodes before the generator pulse arrives at the switch. During conduction, the switch behaves essentially like a short-circuit, completely isolating the load from the generator. At a current

level which is determined by the system characteristics, the plasma ceases to conduct (the switch opens) and the current is rapidly diverted to the load. The conduction and opening stages of the switch operation will be discussed separately.

It has been shown both experimentally⁸ and theoretically⁹ that a sufficiently large current flowing between a plasma and an electrode causes a sheath to develop at the electrode surface. Electron current i and ion current I flow across this sheath. It can be shown from Poisson's equation and conservation of energy that the bipolar current flow is characterized, under space-charge-limited conditions, by the ratio

$$\frac{I}{i} = \left(\frac{Zm}{M} \right)^{1/2} \left(1 + \frac{e\Phi}{2mc^2} \right)^{1/2} = \left(\frac{Zm}{M} \right)^{1/2} \left[\frac{1}{2}(\gamma + 1) \right]^{1/2} \quad (3.1)$$

where m , M are the masses of the electron and ion, respectively, e and Ze their charges, Φ the potential difference across the sheath, and γ the ratio of the total electron energy to its rest energy. The ion current density is given by $J = ZeNV$, where N is the ion density in the plasma and V is the average plasma injection velocity. We shall assume, for the moment, that N is both constant and uniform. Initially, the only current flowing across the gap is the current associated with the flow of ions from the injected plasma. Experiment¹⁰ shows that the (axial) penetration of current into the plasma increases with time. We shall assume that there is no radial variation of current. The ion current is given by

$$I(t) = 2\pi R J L(t) \quad (3.2)$$

where R is the radius of the cathode and $L(t)$ is the width of the current channel. Very soon, however, the electrostatic field at the cathode becomes large enough for field emission of electrons to occur. For the C^{++} ions usually used, $I/i \approx \sqrt{2m/M} \approx 0.01$, i.e. the switch current is due almost entirely to electron flow across the gap, and we may write $I_s = I + i \approx i$. The width of the current channel is then given by

$$L(t) = \left(\frac{Zm}{M} \right)^{1/2} \frac{I_s(t)}{2\pi R J} \quad (3.3)$$

(we have used the non-relativistic expression for the bipolar current ratio), and will broaden as the generator current, and hence the switch current I_s , increases. Switch conduction will continue in this fashion until the width of the current channel becomes equal to the full plasma width L_0 , i.e. when the switch current is given by

$$I_0 = 2\pi R L_0 J (M/Zm)^{1/2}; \quad J = ZeNV \quad (3.4)$$

At this point, the conduction phase ends.

Further increases in switch current require that ions be removed from the plasma faster than they are supplied. The plasma boundary therefore recedes from the cathode, and the gap size D increases, at a rate given by

$$J = ZeN \left(V + \frac{dD}{dt} \right)$$

or

$$\frac{dD}{dt} = \frac{J}{ZeN} - V = \frac{I}{ZeN(2\pi RL_0)} - V \quad (3.5)$$

This process is called "erosion", and signals the beginning of the opening of the switch. The electron current density is given by

$$j = \frac{\epsilon_0}{4} \left(\frac{2e}{m} \right)^{1/2} \frac{\Phi^{3/2}}{D^2} g^2(\alpha) ; \quad \alpha = \frac{e\Phi}{2mc^2} \quad (3.6)$$

where

$$g(\alpha) = \int_0^1 \frac{d\xi}{\left[\xi^{1/2} (1 + \alpha\xi)^{1/2} + (1 + \alpha)^{1/2} \{ (1 - \xi)^{1/2} - 1 \} \right]^{1/2}} \quad (3.7)$$

In the non-relativistic limit ($\alpha \rightarrow 0$), the plasma potential Φ is obtained from

$$i = \frac{\epsilon_0}{4} (2\pi RL_0) \left(\frac{2e}{m} \right)^{1/2} \frac{\Phi^{3/2}}{D^2} g^2(0) \approx I_s \Rightarrow \Phi^{3/2} \propto I_s D^2 \quad (3.8)$$

If the voltage Φ drives current through the load, we may say that the switch has begun to open.

It has recently been suggested⁶ that another mechanism, viz. magnetic pressure, may contribute to the opening of the switch during the erosion phase. The equation of motion of the ions

$$MN \frac{d\mathbf{V}}{dt} = ZeN(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

has a radial component

$$\rho \frac{d}{dt} \left(\frac{dD}{dt} \right) = J_{ax} B \quad (3.8)$$

where ρ is the ion mass density, and the radial component of the $\mathbf{E}$ field has been neglected. Thus an axial component J_{ax} of the ion current density might augment the opening of the gap between the plasma and the cathode, especially if the magnetic field is large. Measurements taken during the GAMBLE II experiment (conduction time 60 ns) indicate an average opening rate, during erosion, of 3.3×10^7 cm/sec. The ion current density required to open the gap at this rate by erosion only is $J \approx 0.5$ kA / cm², which is smaller than the ion current density in other short-conduction-time PEOS experiments. Since the magnetic field is also relatively high, magnetic pressure cannot be ruled out as a contributor to the opening. In GAMBLE II experiments in the 1 ms conduction-time range, the average opening rate is 4 cm/ms, requiring an ion current density of between 3 and 11 kA / cm² for opening by erosion alone. Since the measured ion current density has a peak value of around 1 kA / cm², it can be concluded that erosion is not the only opening mechanism in this case.

It should be borne in mind that the bipolar space-charge-limited condition (3.1) refers to one-dimensional current flow. As the switch current increases further, the magnetic field becomes strong enough to cause appreciable bending of the electron orbit in the gap, i.e. the

motion of the electron becomes two-dimensional. Because of the large mass of the ion, its orbit remains essentially one-dimensional. The space-charge distribution is altered, and the current ratio (3.1) is multiplied by a factor which depends on the inverse ratio of the transit times of the ions and electrons. This factor has been estimated¹¹ to be $(2L_o / D)$, so that the current ratio becomes

$$\frac{I}{i} = \left[\frac{2Zm}{M} (\gamma + 1) \right]^{1/2} \frac{L_o}{D} \quad (3.9)$$

The ion current is therefore enhanced, and the gap opens at a much faster rate. Erosion rates as high as 10^8 cm/sec can be obtained².

This "enhanced erosion" phase continues until the gap size becomes greater than the electron Larmor radius, preventing electron conduction between the electrodes. The only remaining switch current is the single-species Child-Langmuir (ion) current, which is typically much less than the generator current. At this point, "magnetic insulation" of the electrons at the cathode has set in, and the switch may be said to be open. The critical current I_c for the onset of magnetic insulation (when the width D of the gap is equal to the Larmor radius of the electron) is given by

$$\gamma m \left(\frac{v^2}{D} \right) = ev \left(\frac{\mu_o I_c}{2\pi R} \right) \quad (3.10)$$

since the thickness of the sheath is much less than the cathode radius. Using the result

$$v = \left[(\gamma^2 - 1)^{1/2} / \gamma \right] c \quad (3.11)$$

we get

$$I_c = \left(\frac{mc}{e} \right) \left(\frac{2\pi}{\mu_o} \right) (\gamma^2 - 1)^{1/2} \left(\frac{R}{D} \right) \quad (3.12)$$

The switch is fully open when the load current I_L exceeds the critical current I_c for magnetic insulation of the electron flow, i.e. when $I_L = \Phi / Z_L > I_c$, where Z_L is the load impedance.

Using (3.12) and the result $\gamma = 1 + (e\Phi / mc^2)$, the critical voltage across the load can then be shown to be

$$\Phi_c = \frac{(0.16 Z_L R / D)^2}{1 - (0.16 Z_L R / D)^2} \times 10^6 \text{ volts} \quad (3.13)$$

We now consider the system requirements for optimum switch performance. As can be seen from Equation (3.4), for given Z and cathode surface $A = 2\pi RL_o$, the maximum switch current before the erosion phase begins is proportional to NV , where the ion velocity V is taken to be the plasma injection velocity V_d . The erosion rate is given by

$$\frac{dD}{dt} = \frac{J}{ZeN} - V_d = \frac{I}{ZeN(2\pi RL_o)} - V_d \quad (3.14)$$

Thus if $V_d \ll \frac{dD}{dt}$, the erosion rate is inversely proportional to the ion density N , for fixed R , L_o , Z and I . A high-velocity, low-density plasma should therefore be used in order to achieve both high conduction current and fast opening. Typical values are plasma injection velocity $V_d \approx 10^7$ cm/sec and doubly ionized Carbon density $N \approx 10^{13}$ cm⁻³. For complete magnetic insulation of the electrons at the cathode, the load voltage must be greater than the critical value given in (3.13).

IV. DYNAMICS OF CURRENT FLOW IN THE PLASMA

The four-phase theory successfully describes the gross electrical behavior of the PEOS. Gap formation has been shown to progress axially along the cathode during the conduction phase, with opening proceeding radially via erosion (and possibly magnetic pressure). The (maximum) conduction current is proportional to the plasma density, injection velocity and cathode area, while the erosion rate is (approximately) inversely proportional to the ion density. Conspicuously absent from this analysis, however, is a discussion of such matters as the role of the magnetic field produced by the switch current and the dynamics of current flow in the plasma. In a number of experiments¹⁰, arrays of magnetic probes have been used to map the spatial and temporal distribution of magnetic field in the plasma during the conduction phase. These measurements show that the magnetic field penetrates axially into the plasma, but that the current channel width is less than the field penetration into the plasma. The dynamics of current flow in the plasma has been described in terms of a two-fluid model, which encompasses the hydrodynamics of the plasma, Maxwell's equations and a two-dimensional electron orbit analysis. The purpose of this approach is to investigate the role played by two mechanisms, viz. axial magneto hydrodynamic compression of the plasma due to $\mathbf{J} \times \mathbf{B}$ forces and magnetic insulation of electrons emitted from the cathode, in the conduction process.

This fluid simulation of the conduction phase of the PEOS uses a Cartesian system of coordinates (which approximates the experimental geometry) in which the x -axis points in the (outward) radial direction, the y -axis in the azimuthal direction and the z -axis in the axial direction (towards the load). Lower case letters will be used to denote the electron variables (mass m , density n , velocity $\mathbf{v}$, etc.), with upper case letters (M , N , V , etc.) for the

corresponding ion variables. It is assumed that all of the field and fluid variables depend on the single spatial coordinate z and the time t , i.e. $\frac{\partial}{\partial x} = \frac{\partial}{\partial y} = 0$. The magnetic field is taken to be in the (negative) azimuthal direction ($\mathbf{B} = -B \hat{y}$).

We start with the conservation of the number of electrons, which is represented by the continuity equation

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial z}(nv_z) = 0 \quad (4.1)$$

The equation of motion is

$$m \frac{d\mathbf{v}}{dt} + \nabla p = -e (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (4.2)$$

where $\frac{d}{dt} = \frac{\partial}{\partial t} + v_z \frac{\partial}{\partial z}$ is the convective derivative and p is the (scalar) pressure. We shall neglect the pressure term for the moment. The x -component of (4.2) is then

$$m \left[\frac{\partial v_x}{\partial t} + v_z \frac{\partial v_x}{\partial z} \right] = -e [E_x + v_z B] \quad (4.3)$$

This may be written in terms of the electron momentum density by multiplying (4.3) by n , (4.1) by mv_x and adding the results. This gives

$$\frac{\partial}{\partial t}(nmv_x) + \frac{\partial}{\partial z}(v_z nmv_x) = -ne(E_x + v_z B) \quad (4.4)$$

For the z -component, we have

$$\frac{\partial}{\partial t}(nmv_z) + \frac{\partial}{\partial z}(v_z nmv_z) = -ne(E_z - v_x B) \quad (4.5)$$

A similar set of equations is obtained for the ions:

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial z}(NV_z) = 0 \quad (4.6)$$

$$\frac{\partial}{\partial t}(NMV_z) + \frac{\partial}{\partial z}(V_z NMV_z) = ZeN(E_z - V_x B) \quad (4.7)$$

The radial component of the ion velocity is assumed to be constant, and equal to the plasma injection velocity, i.e.

$$V_x = -V_d \quad (4.8)$$

where V_d is the injection speed. We now turn to Maxwell's equations. Faraday's Law takes the form

$$\frac{\partial E_z}{\partial z} = \frac{\partial B}{\partial t} \quad (4.9)$$

Gauss' Law becomes

$$\frac{\partial E_z}{\partial z} = \frac{1}{\epsilon_0} (ZeN - en) \quad (4.10)$$

and Ampere's Law gives

$$\frac{\partial E_z}{\partial t} = c^2 \frac{\partial B}{\partial z} - \frac{1}{\epsilon_0} [ZeNV_z - nev_z] \quad (4.11)$$

The conduction phase of the PEOS has been studied with the aid of an electromagnetic two-fluid code which is based on the system of 8 equations (4.1), (4.4) - (4.7) and (4.9) - (4.11). The Maxwell's equations are solved numerically, the fluid equations are solved using a flux-corrected-transport (FCT) algorithm, and the cathode sheath physics is described by using the bipolar, space-charge-limited condition (3.1). The eight dependent variables N , n , V_z , v_z , B , E_z , E_r are determined by application of the appropriate boundary conditions. Two-dimensional electron orbits are calculated by injecting test electrons into the electromagnetic fields derived by the fluid code.

Alternatively, the degree of compression of the plasma can be described in terms of a one-dimensional treatment of the ion hydrodynamics. The following simplifying assumptions are made:

- (i) The velocity u at which the current channel penetrates the plasma is constant.
- (ii) The plasma is quasi-neutral, i.e. the electron density n and the ion density N are related by $en \approx Zen$, or $n \approx ZN$.
- (iii) All of the fluid and field variables depend on the space coordinate z and the time t through the single variable $\xi = ut - z$.
- (iv) The current is carried predominantly radially, by $\mathbf{E} \times \mathbf{B}$ drift of electrons.
- (v) The electron current density satisfies the bipolar space-charge-limited condition (3.1).

Conservation of the number of ions is expressed by the continuity equation

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial z} (NV_z) = 0 \quad (4.6)$$

The equation of motion is

$$MN \left[\frac{\partial V_z}{\partial t} + V_z \frac{\partial V_z}{\partial z} \right] + \frac{\partial}{\partial z} \psi_a = ZeNE_z \quad (4.12)$$

where the term BV_z has been neglected in comparison with E_z . The electron current density is given by

$$j = en \left| \frac{\mathbf{E} \times \mathbf{B}}{B^2} \right| = ZeN \frac{E_z}{B} \quad (4.13)$$

using the assumptions of quasi-neutrality and radial current density (which implies that $E_r = 0$). Equations (4.12) and (4.13) can be combined to give

$$MN \left(\frac{\partial V_z}{\partial t} + V_z \frac{\partial V_z}{\partial z} \right) + \frac{\partial}{\partial z} \psi_a = jB \quad (4.14)$$

The displacement current density may be neglected in comparison with the massive electron current density. Thus Ampere's law may be written

$$\frac{\partial B}{\partial z} = -\mu_0 j \quad (4.15)$$

The quantities of immediate interest are the ion density N , the axial ion flow velocity V_z and the plasma potential Φ . These may be obtained from the equations (4.6), (4.14) and (4.15) together with the fields derived from the fluid code and the bipolar space-charge-limited condition. Let us recall that all variables are assumed to depend on the single variable $\xi = ut - z$, where u is constant. Thus

$$\frac{\partial}{\partial t} = \frac{\partial \xi}{\partial t} \frac{d}{d\xi} = u \frac{d}{d\xi} ; \quad \frac{\partial}{\partial z} = \frac{\partial \xi}{\partial z} \frac{d}{d\xi} = -\frac{d}{d\xi}$$

and Equation (4.6) becomes

$$-\frac{d}{d\xi} (NV_z) + u \frac{dN}{d\xi} = 0 \quad (4.16)$$

The boundary of the undisturbed plasma is $u = 0$, so the relevant boundary conditions are: at $u = 0$, $B = 0, N = N_0$ (the original ion density), $V_z = 0$ and $\Phi = 0$. Thus, (4.16) may be integrated to give

$$NV_z = (N - N_0) u$$

$$V_z = \left(1 - \frac{N_0}{N} \right) u \quad (4.17)$$

Equations (4.14) and (4.15) may be combined to give

$$\mu_o MN(u - V_z) \frac{dV_z}{d\xi} - \mu_o \frac{d}{d\xi} \psi_z = \frac{1}{2} \frac{d}{d\xi} B^2 \quad (4.18)$$

which, upon substituting (4.17) into (4.18) and integrating, yields

$$2\mu_o MN_o^2 u^2 \int_{N_o}^N \frac{dN}{N^2} = \int_0^B d(B^2 + 2\mu_o \psi_z) \quad (4.19)$$

whence

$$B_o^2 \left(1 - \frac{N_o}{N}\right) = B^2 + 2\mu_o \psi_z$$

or

$$N = \left[1 - \frac{B^2 + 2\mu_o \psi_z}{B_o^2}\right]^{-1} N_o \quad (4.20)$$

where

$$B_o^2 = 2\mu_o MN_o u^2 \quad (4.21)$$

Equation (4.20) gives the hydrodynamic compression of the plasma in terms of the velocity u , the magnetic field B and the pressure tensor component ψ_z . It is difficult to determine ψ_z , but its exclusion from the analysis, which yields the approximate formula

$$N = \left[\frac{1}{1 - (B/B_o)^2} \right] N_o \quad (4.22)$$

may lead to unphysical plasma compression (for $B \geq B_o$)

The plasma potential $\Phi = -\int E_z dz$ is obtained by integrating Equation (4.12) to give

$$\frac{Ze}{M} \Phi = \int \left[u \frac{dV_z}{d\xi} - V_z \frac{dV_z}{d\xi} - \frac{1}{N} \frac{d\psi_z}{d\xi} \right] d\xi \quad (4.23)$$

$$\Phi = \frac{Mu^2}{Ze} \left[\left(\frac{B'}{B_o} \right)^2 - \frac{1}{2} \left(\frac{B'}{B_o} \right)^4 - \frac{1}{u^2 N} \psi_z \right] \quad (4.24)$$

where N is given in terms of V_z by Equation (4.17). The pressure tensor component ψ_z is, however, not known. An approximate expression for Φ has been obtained by neglecting ψ_z in (4.24) altogether. This yields the result

$$\Phi = \frac{Mu^2}{Ze} \left[\left(\frac{B}{B_o} \right)^2 - \frac{1}{2} \left(\frac{B}{B_o} \right)^4 \right] \quad (4.25)$$

Electrons emitted from the cathode are accelerated across the cathode sheath and injected into the plasma with energies corresponding to this potential. Variation of the B field with ξ (i.e. penetration of the B field into the plasma) is obtained from

$$\frac{\partial B}{\partial z} = -\mu_o j \Rightarrow -\frac{\partial B}{\partial \xi} = -\mu_o \left(\frac{M}{Zm} \right)^{1/2} ZeNV_d$$

or

$$\frac{\partial B}{\partial \xi} = \mu_o \left(\frac{M}{Zm} \right)^{1/2} \left[\frac{N_o Ze V_d}{1 - (B/B_o)^2} \right] \quad (4.26)$$

where the pressure term has again been neglected. This equation may be integrated to give

$$\xi = \left[\left(\frac{B}{B_o} \right) - \frac{1}{3} \left(\frac{B}{B_o} \right)^3 \right] \xi_H \quad (4.27)$$

where

$$\xi_H = \frac{2mu^2}{\mu_o N_o Ze^2 V_d^2} \quad (4.28)$$

Only the dynamics of ion flow in the plasma has been considered so far. We must now turn our attention to the emission of electrons from the cathode, their migration across the cathode sheath and subsequent injection into the plasma. The situation has been investigated⁶ for the case of non-relativistic electrons injected into electric and magnetic fields which are assumed to be constant and uniform. This analysis shows that some of these electrons drift radially across the plasma, but others describe orbits which eventually return them to the cathode, i.e. they are magnetically insulated. The orbit equations are

$$x(t) = \frac{a}{\omega} [\sin(\omega t - \delta) + \sin \delta] + \frac{E}{B} t \quad (4.29)$$

$$z(t) = -\frac{a}{\omega} [\cos(\omega t - \delta) - \cos \delta] + z_o \quad (4.30)$$

where $\omega = eB/m$ and the initial conditions $x(0) = 0, z(0) = z_o$ have been used. For convenience, we shall write $a = k(E/B)$ and normalize the electron velocity to E/B . The normalized injection velocity (i.e. at time $t = 0$) is given by

$$\beta = \frac{v_{inj}}{(E/B)} = (k^2 + 2k \cos \delta + 1)^{1/2} \quad (4.31)$$

and the injection angle Θ_o , measured with respect to the positive z -axis, is

$$\Theta_o = \begin{cases} \pi - \tan^{-1}[(k \cos \delta + 1) / k \sin \delta] ; 0 \leq \delta \leq \pi \\ \tan^{-1}[-(k \cos \delta + 1) / k \sin \delta] ; \pi \leq \delta \leq 2\pi \end{cases} \quad (4.32)$$

The condition that an electron, once injected into the plasma, should not return to the cathode (i.e. be "lifted off") is

$$x(t) > 0, \forall t$$

or

$$t \left[\frac{k}{\omega t} \{ \sin(\omega t - \delta) + \sin \delta \} + 1 \right] > 0, \forall t \quad (4.33)$$

Since $t > 0$, this means that

$$k \left\{ \frac{\sin \delta - \sin(\delta - \theta)}{\theta} \right\} + 1 > 0 \quad (4.34)$$

i.e. $k g(\theta) > -1 (\theta = \omega t > 0)$,

where

$$g(\theta) = \frac{\sin \delta - \sin(\delta - \theta)}{\theta} \quad (4.35)$$

Let $g(\alpha) = -M$ be the minimum value of g . Clearly, we must have $k g(\alpha) > -1$, i.e. $k < k_M$, where

$$k_M = -\frac{1}{g(\alpha)} = -\frac{\alpha}{\sin \delta - \sin(\delta - \alpha)} \quad (4.36)$$

or

$$k_M \sin \delta - k_M \sin(\delta - \alpha) = -\alpha \quad (4.37)$$

From the requirement $g'(\alpha) = 0$, it is easy to see that

$$\cos(\delta - \alpha) = \frac{\sin \delta - \sin(\delta - \alpha)}{\alpha} = -\frac{1}{k_M} \quad (4.38)$$

Substituting (4.38) into (4.36), we get

$$k_M \sin \delta - \sqrt{k_M^2 - 1} + \delta - \cos^{-1}(-1/k_M) + 2\pi = 0 \quad (4.39)$$

For a given phase angle δ , Equation (4.39) yields the critical value k_M , which in turn gives the critical (normalized) injection velocity β_c , using Equation (4.31). The angle of injection Θ_o is then obtained from (4.32). For normal incidence ($\Theta_o = \pi/2$), $\beta_c \approx 5.6$, i.e. electrons injected

normally will lift off immediately if their (normalized) injection velocity $\beta < 5.6$. Electrons injected with greater velocities will experience magnetic insulation. We mention, in passing, that β can be related to the switch parameters, as follows:

$$\beta = \frac{v_{inj}}{(E/B)} = \frac{(2e\Phi/m)^{1/2}}{j/ne} \quad (4.40)$$

Now using the bipolar space-charge-limited condition and assuming quasi-neutrality ($ne = ZeN$), we get

$$\beta = \frac{(2e\Phi/m)^{1/2}}{(M/Zm)^{1/2}V_d} = \left(\frac{2Ze}{M}\right)^{1/2} \frac{\Phi^{1/2}}{V_d} \quad (4.41)$$

But from (4.23) it follows that $\Phi \leq \frac{1}{2} \left(\frac{Mu^2}{Ze} \right)$. Hence we obtain

$$\beta \leq \frac{u}{V_d} \quad (4.42)$$

where u is the velocity of the current channel front and V_d is the plasma injection velocity. The criterion for immediate lift-off of electrons from the cathode is $\beta < \beta_c$, and this will be automatically satisfied if the ratio u/V_d is less than β_c . It is therefore desirable to have large plasma injection speeds, a condition which also favors high conduction current before the onset of erosion and rapid switch opening afterwards.

The effect of magnetic insulation in limiting the width of the current channel may be investigated by calculating the orbits of test electrons injected into the fields derived from the fluid code. In the analysis carried out by Grossmann et al⁶, a number of simplifying assumptions have been made: non-relativistic orbit equations have been used, the loss of energy of the electrons as they are decelerated by the E_z field has been neglected and the effect of the cathode sheath has been ignored, except insofar as it serves to determine the initial injection velocity of the electrons. Electrons which fail to lift-off immediately describe orbits which return them to the cathode, where they undergo specular reflection and are returned to the system at a larger injection angle. This process continues until the injection angle becomes so large that the injection speed is less than the critical velocity for electron lift-off. This critical velocity for electron lift-off increases rapidly with injection angle Φ_0 for $\Phi_0 \geq \pi/2$, which clearly facilitates eventual lift-off.

V. DISCUSSION

In this report, we have attempted a review of the progress that has been made so far in understanding the basic principles underlying the operation of the Plasma Erosion Opening Switch. A theory has been developed at NRL in which the switch operation is envisaged as a sequence of four distinct phases: conduction, erosion, enhanced erosion and magnetic insulation. While this four-phase theory describes the gross behavior of the switch, it fails to deal with such important matters as cathode emission physics and the dynamics of current flow in the plasma. Grossmann et al⁶ have developed a two-fluid electromagnetic code with these deficiencies in mind. As a parallel undertaking, they have carried out a theoretical analysis of the motion of the ions in the plasma, using the fields derived from the fluid code. As a result, they have identified two mechanisms (plasma compression and magnetic insulation of electrons at the cathode) which play an important role in the conduction phase. Unfortunately, the values of the current channel width obtained by the fluid code and the analytical approach do not agree with each other or with experimental observations. However, it is found that if the injection energies of the electrons are reduced, the effects of magnetic insulation are decreased, and better agreement with experiment is achieved. This lends credence to the view that one or more energy loss mechanisms may come into play in the vicinity of the cathode.

It must be emphasized that the review presented here is far from complete. Time and space constraints have permitted only a limited discussion of work in progress in this area. More detailed analyses^{12,13} of the conduction phase have recently been made, and these must be included in order to arrive at a fuller understanding of the operation of the switch.

REFERENCES

1. P. F. Ottinger, S. A. Goldstein and R. A. Meger, "Theoretical Modelling of the Plasma Erosion Opening Switch for Inductive Storage Applications", J. Appl. Phys. 56, 224 (1984).
2. R. J. Commisso, G. Cooperstein, R. A. Meger, J. M. Neri, P. F. Ottinger and B. V. Weber, "The Plasma Erosion Opening Switch", NRL Memorandum Report 5560 (1985).
3. G. Cooperstein and P. F. Ottinger, "Guest Editorial: Fast Opening Vacuum Switches for High-Power Inductive Energy Storage", IEEE Trans. Plasma Sci., PS-15, No. 6, 629 (1987).
4. B. V. Weber, R. J. Commisso, G. Cooperstein, J. M. Grossmann, D. D. Hinshelwood, M. Mosher, J. M. Neri, P. F. Ottinger and S. J. Stephanakis, "Plasma

Erosion Opening Switch Research at NRL", IEEE Trans. Plasma Sci. PS-15, No. 6, 635 (1987).

5. B. V. Weber, R. J. Commisso, P. J. Goodrich, J. M. Grossmann, D. D. Hinshelwood, J. C. Kellogg and P. F. Ottinger, "Investigation of Plasma Opening Switch Conduction and Opening Mechanisms", IEEE Trans. Plasma Sci., PS-19, No. 5, 757 (1991).
6. J. M. Grossmann, D. Mosher and P. F. Ottinger, "Fluid Simulation of the Conduction Phase of the Plasma Erosion Opening Switch", IEEE Trans. Plasma Sci., PS-15, No. 6, 704 (1987).
7. D. Mosher, J. M. Grossmann, P. F. Ottinger and D. G. Colombant, "A Self-Similar Model for Conduction in the Plasma Erosion Opening Switch", IEEE Trans. Plasma Sci., PS-15, No. 6, 695 (1987).
8. P. A. Miller, J. W. Poukey and T. P. Wright, Phys. Rev. Lett., 35, 940 (1975).
9. A. G. Jack, K. F. Sander and R. H. Valey, J. Plasma Phys., 5, 211 (1971).
10. B. V. Weber, R. J. Commisso, R. A. Meger, J. M. Neri, W. F. Oliphant and P. F. Ottinger, "Current Distribution in a Plasma Erosion Opening Switch", Appl. Phys. Lett., 45(10), 1043 (1984).
11. S. A. Goldstein and R. Lee, "Ion-induced Pinch and the Enhancement of Ion Current by Pinched Electron Flow in Relativistic Diodes", Phys. Rev. Lett., 35(16), 1079 (1975).
12. Special Issue, IEEE Trans. Plasma Sci., PS-15, No. 6, (1987).
13. J. M. Grossmann, J. M. Neri and P. F. Ottinger, "Investigation of Collisional Effects in the Plasma Erosion Opening Switch", NRL Memorandum Report 6240 (1988).

ACKNOWLEDGMENTS

I would like to express my gratitude to the Air Force Office of Scientific Research for sponsoring this Research and to the Arno'd Engineering Development Center for its hospitality during the Summer. In particular, I would like to thank Mr. Lavell Whitehead for suggesting this topic and the many AEDC personnel who have been so helpful at all times. Lastly, I wish to thank Captain Hal Martin of the USAF for his oversight of the Program and the staff at Research and Development Laboratories (RDL) for their cheerful and efficient administration.

**EXTENSION OF THE SMIRF FLARE MODEL TO
COMPLEX GEOMETRICAL SHAPES**

**Olin Perry Norton
Research Engineer
Diagnostic Instrumentation and Analysis Laboratory**

**Mississippi State University
P. O. Drawer MM
Mississippi State, MS 39762**

**Final Report for:
Summer Research Program
Arnold Engineering Development Center**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.**

September 1992

EXTENSION OF THE SMIRF FLARE MODEL TO COMPLEX GEOMETRICAL SHAPES

Olin Perry Norton
Research Engineer
Diagnostic Instrumentation and Analysis Laboratory
Mississippi State University

Abstract

The SMIRF flare model was written to calculate the infrared signature of flares. The existing code was limited to flares whose shapes were simple geometric forms, such as cylinders, spheres, and rectangular parallelepipeds. This work describes the removal of that restriction, and the modifications that were made to the code to allow for more complex geometrical shapes.

The code was modified to read the geometrical data from a user-supplied data file. If an appropriate data file can be supplied, then SMIRF will be able to compute the signature for flares of arbitrary shape. This data file contains information describing how the surface area and volume of the flare change as the flare burns.

In addition, an approximation method is developed for modeling flares with grooved surfaces. This approximation expresses the increased surface area of the grooved surface in terms of an area augmentation ratio. Thus, the area of the surface is calculated ignoring the grooves, and the result is multiplied by the augmentation ratio to find the area of the grooved surface.

EXTENSION OF THE SMIRF FLARE MODEL TO COMPLEX GEOMETRICAL SHAPES

Olin Perry Norton

INTRODUCTION

The acronym SMIRF stands for Signature Model for InfraRed Flares. The SMIRF computer program has been under development for several years at Arnold Engineering Development Center (AEDC), where the author of this report was a participant in the summer faculty research program. SMIRF is the first program to attempt to predict the infrared signatures of flares from first principles. Additional information on this code can be found by consulting references [1], [2], and [3].

Flares are pyrotechnic devices, which burn and produce intense heat. The flares of interest here are intended as infrared decoys. The burning flare produces hot combustion products, which in turn produce infrared radiation. The infrared radiation is supposed to lure an infrared seeking missile away from the aircraft it is tracking. Thus, the infrared output of the flare is an important indicator of its potential effectiveness. With this in mind, the importance of the SMIRF model and its ability to predict the infrared signature is readily apparent.

Infrared decoy flares are usually composed of powdered magnesium mixed with polytetrafluoroethylene (PTFE, better known as Teflon®). This mixture is held together with a plastic binder and is pressed or molded into a solid grain. When ignited, the PTFE and magnesium react vigorously. This reaction is very similar to the burning of a solid rocket propellant in that the burning occurs only at the exposed surface of the solid; the burning rate of the flare will be proportional to the surface area of the flare grain. This gives the flare designer some control over the flare signature, since the grain shape can be altered to provide a greater surface area and thus a faster burning rate. As a consequence, some flares have complicated geometrical shapes, with grooved surfaces and internal holes.

Clearly, since these geometrical features increase the surface area of the flare, these features will increase the burning rate and thus the infrared output of the flare. If SMIRF is to be successful in predicting the signature of these flares, then SMIRF must have the ability to predict the burning rates of these geometrically complex shapes.

The existing SMIRF model was limited to flare grains which were either rectangular parallelepipeds, cylinders, or spheres. One of the tasks which the author performed during the summer of 1992 was to extend the capability of SMIRF to handle more realistic grain shapes.

SMIRF DEPENDENCE ON FLARE GEOMETRY

To understand how the SMIRF model can be modified to handle oddly shaped flares, it is first necessary to understand some details about how SMIRF works, the assumptions that are made in SMIRF, and how the flare shape enters into the calculations.

First, SMIRF finds the flare trajectory. Newton's laws of motion are integrated numerically to find the position and velocity of the flare as functions of time after the flare is ejected from an aircraft. Since the flare is burning as it falls, the mass of the flare is decreasing. The rate of decrease in the flare mass is equal to the burning rate of the flare, defined as the rate at which the solid flare material is being converted into combustion products. The burning rate typically has units of pounds mass per second. SMIRF must correctly model the burning rate to compute the trajectory.

The trajectory calculation includes the effects of aerodynamic drag. The drag model incorporated in SMIRF does not depend on the geometrical details, but models the drag in terms of the volume and the surface area of the flare. Therefore, the flare surface area and volume must be provided in order to correctly compute the aerodynamic drag, and thus the trajectory.

Once the trajectory of the flare is found, the flow field in the flare plume is modeled using the standard JANNAF plume model, SPF-2. SPF-2 was intended to model rocket plumes and assumes that the flow field is axially symmetric [4]. As a result, SMIRF must ignore the geometric details of the flare shape and replace the flare with an axially symmetric shape that is somehow equivalent to the actual flare shape.

Defining equivalence in this context depends on identifying the characteristics of the flare that are most important in affecting the infrared signature. The SMIRF flare model calculates the infrared radiation from the plume formed as the flare products mix with the free stream, and it is these combustion products that are responsible for most of the flare signature. Therefore, the most important characteristics to preserve when defining an equivalent axisymmetric flare shape are the velocity and mass flow rate of the combustion products.

SMIRF models the flare as a sphere that is burning and ejecting hot combustion products. The sphere diameter is determined such that the surface area of this sphere is the same as the surface area of the actual flare. This is done so that the mass flux and velocity of combustion products into the plume will be the same as for the actual flare. Conservation of mass requires that the mass flow rate of combustion products be equal to the rate of burning of the flare.

To summarize, SMIRF calculations do not depend on the exact details of the flare shape, but instead depend on certain gross features of the flare: the volume, the surface area, and the mass flux of combustion products from the flare (the burning rate). Modifying SMIRF for different flare shapes is simply a matter of correctly specifying these quantities. The volume and surface area are clearly defined geometrical properties, and the burning rate also depends on the geometry of the flare.

BURNING RATE CALCULATIONS

These flares are solid objects that burn at the surface, and the calculation of the total burning rate is very similar to the procedure for solid propellant rockets. The greater the exposed surface area, the faster the flare burns.

Letting $\dot{m}$ represent the rate of burning of the flare, it is clear that this term is the product of three factors:

$$\dot{m} = \beta \rho_s A_b \quad (1)$$

where β is the burning velocity of the solid pyrotechnic material that composes the flare (typically about 0.1 inches/second), ρ_s is the density of the solid flare material (typically about 0.05 pounds/cubic inch) and A_b is the area of the burning surface of the flare (measured in square inches). In most cases, A_b will be the entire external surface area of the flare grain, but, in cases where only part of the surface is burning, only the part of the surface that is burning is included.

The determination of β is a weak spot in the calculation of the flare signature from basic principles. The user of the SMIRF code can specify a constant value for β or a power-law dependence on pressure can be specified. In either case, the user must have some experimental data in hand for the particular grain composition. In defense of SMIRF's developers, it should be pointed out that this is essentially the same problem as predicting the burning velocity of a solid rocket propellant, and this author is unaware of any success in predicting this from first principles.

The density of the solid flare material will be known. In most cases, the volume and mass of the unburned flare will be given, and the density can be found from these quantities. The prediction of burning rate is thus mainly a matter of determining A_b . As the flare is consumed, the surface area of the grain changes causing the flare output to change. Eventually, this surface area must go to zero, corresponding to the complete consumption of the solid grain and burnout. Thus, A_b will vary as the flare burns and the accurate prediction of the flare output depends on accurately modeling this variation.

The existing SMIRF flare model contains explicit algebraic equations for the surface area and volume of the flare for some simple geometric shapes. The shapes included in the model are:

- Rectangular parallelepiped (overall burning)
- Right circular cylinder (overall burning)
- Right circular cylinder (end burning)
- Sphere (overall burning)

The SMIRF user specifies one of these geometries and provides the initial dimensions of the flare. The SMIRF program then calculates the varying surface area and volume of the flare as it burns.

As an example consider the first case, a flare in the shape of a rectangular parallelepiped. (This is a "brick" shaped flare.) Suppose the initial dimensions of the flare are X by Y by Z . Assuming a constant β , after the flare has burned for a time t , the flame front will have burned a distance $\delta = \beta t$. Each face of the parallelepiped will have moved inward a distance δ , so the surface area of flare will be:

$$A_b = 2 ((X - 2\delta)(Y - 2\delta) + (X - 2\delta)(Z - 2\delta) + (Y - 2\delta)(Z - 2\delta)) \quad (2)$$

For the trajectory calculation, SMIRF also needs to know the mass of the solid flare at each instant of time. This can be found from the flare solid density and the instantaneous volume of the flare. For this example, the volume will be:

$$V = (X - 2\delta)(Y - 2\delta)(Z - 2\delta) \quad (3)$$

These equations will be valid until the flare burns out. For example, if Z is the smallest initial dimension of the flare, these equations will hold as long as $(Z - 2\delta)$ is nonnegative.

The above equations, stated in terms of the flame burn distance δ , will still hold even if the burning velocity β is not constant. For time dependent β , the burn distance is found by integrating:

$$\delta = \int_0^t \beta(t) dt \quad (4)$$

and equations (2) and (3) above will still be true.

The determination of the burning area and volume are straightforward for the simple geometrical shapes listed above. But, not all flare grains fall into one of these categories. The surface of the flare may be grooved to increase the burning surface area and the flare may contain internal holes. These modifications will increase

the burning surface area of the flare and thus the infrared output. Accordingly, the SMIRF flare model was modified to allow for nonstandard shapes.

AERODYNAMIC SURFACE AREA AND VOLUME

Generalizing the SMIRF flare model to include complex shapes with grooved surfaces and internal holes requires that two additional geometrical quantities be defined. These quantities are the aerodynamic surface area and the aerodynamic volume.

The existing SMIRF flare model uses the flare surface area in two different places in the calculations. The first use of the surface area, as discussed above, is to find the burning rate of the flare. For this purpose, it is appropriate to include the surface area due to surface irregularities (such as grooves) and the additional area due to internal holes in the flare. However, in a case such as an end burning cylinder ("cigar"), only the actual burning surface should be included.

SMIRF also uses the flare surface area in the trajectory calculation to find the aerodynamic drag on the flare. For this purpose, the surface area should include non-burning surfaces, but probably should not include the surface area added by internal passages and grooves.

Thus, complex shapes require that two different surface areas be defined. The first, A_b , the burning surface area, is used to find the rate of burning of the flare. To find the aerodynamic drag of the flare, a second surface area, A_d , is defined. This can be called the aerodynamic surface area.

Similarly, SMIRF uses the flare volume in two different contexts. One use, as described above, was to find the instantaneous mass of the flare. The other use was to find the aerodynamic drag. Again, internal passages and voids do reduce the mass, but don't change the aerodynamic drag, so two different volumes are needed for these two different purposes. Therefore, an additional quantity, the aerodynamic volume V_d , is introduced.

SMIRF can then be used to compute the signature of a flare of arbitrary geometrical shape provided that these four quantities: burning area, aerodynamic area, volume, and aerodynamic volume are somehow provided. Thus, extending the model to complex shapes is a matter of specifying these areas and volumes at different times in the lifetime of the flare.

This approach is practical because these areas and volumes are purely geometrical quantities which are functions only of δ . For example, consider some arbitrarily shaped solid body. Suppose that a value of the burn distance δ is specified. Start at the surface of the body, and, propagating normal to the surface, move inward by a distance δ . This action defines a new shape. Given this new shape, the surface areas and volumes can be found.

Given some initial shape; the burning area, aerodynamic area, volume, and aerodynamic volume can be found as functions of the burn distance δ . This calculation involves only geometry; starting at the original surface, moving inward a distance δ , and then calculating the areas and volumes of the resulting shape. Once the burning area, aerodynamic area, volume, and aerodynamic volume are determined as functions of the burn distance, SMIRF will be able to compute the trajectory and signature.

DETAILED DESCRIPTION OF SMIRF MODIFICATIONS

Modifying SMIRF to make these geometric calculations for an arbitrary shaped flare would be a tall order. Even ignoring the question of how to do the calculations, it would be difficult to devise an input format that would allow the user to specify an arbitrary initial flare shape. Thus, it was decided to separate SMIRF from the area and volume calculations.

The plan is this: The user takes the initial flare grain shape and computes the burning area, aerodynamic area, volume, and aerodynamic volume as functions of δ . The results of these calculations are tabulated and provided to SMIRF in a data file. When SMIRF needs to know the burning area, aerodynamic area, volume, or aerodynamic volume, SMIRF will compute the value of δ and then perform a table look up and interpolation procedure. This plan was implemented in the following steps:

The existing SMIRF flare model does not make the distinction between the burning area and the aerodynamic area, or between the volume and the aerodynamic volume. The first step in modifying the code was to introduce additional variables for the aerodynamic quantities. Then, each time the code referenced the flare area or volume, this reference was changed as necessary to refer to the proper quantity. The aerodynamic area and aerodynamic volume were used for finding the aerodynamic drag of the flare. The volume was used whenever the flare mass was calculated, and the burning area was used whenever the flare burning rate was calculated.

An additional logical variable was added to the main input file for SMIRF. When this variable is "TRUE," then the SMIRF model will bypass the algebraic equations for simple geometrical shapes and will instead expect to find an additional input file named "geo.dat." In this input file, the variation of area and volume are tabulated. SMIRF will read this data. Then, whenever SMIRF needs to know the area or volume of the flare, SMIRF will look up the necessary data in the table and then interpolate.

The first step in this process is finding the right location in the table. Because SMIRF is numerically integrating the trajectory equations, the subroutine that is called to perform the table look up function is called many times. Thus, it is important to find the right table location efficiently.

The table look up problem is, given a value of δ , and a set of tabulated values δ_i , ordered such that $\delta_{i+1} > \delta_i$, find the value of n such that $\delta_n < \delta < \delta_{n+1}$. This would be easy if the δ_i were evenly spaced. But, to allow the user more flexibility, this assumption was not made and another scheme was implemented instead.

Because a numerical differential equation solver is used to find the trajectory, the table look up subroutine will be called often. Usually the value of δ will be very close to its value the previous time the subroutine was called, and will often correspond to the same value of n . Thus, the table look up subroutine first checks to see if the value of n used the previous time will work again. If not, an efficient patterned search scheme, described by Press, et al. [5], is used to find n . Once the correct value of n is found, linear interpolation is used in the interval from δ_n to δ_{n+1} to find the burning surface area, aerodynamic area, volume, and aerodynamic volume.

The user must provide a file called "geo.dat" which contains all the necessary geometric data. This file is an ASCII formatted data file. Each record in the file should contain five numbers: the burn distance δ (in inches), the burning surface area A_b (in square inches), the aerodynamic surface area A_A (in square inches), the flare volume V (in cubic inches), and the aerodynamic volume V_A (in cubic inches). The input format used for reading "geo.dat" is a freeform FORTRAN read statement; these five values do not have to be in any particular columns, but must be separated by tabs, commas, or one or more blank spaces.

When creating "geo.dat," it is important to understand that the burn distance, δ , in the first column is the independent variable, and the values in the next four columns are the dependent variables which are functions of δ . It is necessary that this data be provided for values of δ from zero (the initial flare shape) to burnout (where the areas and volumes are zero). The table must be ordered so that the value of δ on each line is greater than the value on the line before. However, the increment does not need to be constant.

At present, the maximum number of lines allowed in the file "geo.dat" is 500. This limit is established by the size chosen for the arrays that store the tables. If this number needs to be increased, the value of the parameter NTMAX should be increased in the subroutines READ_TABLE and GET_TABLE. Note that the value of NTMAX should be the same in both places; otherwise a misalignment problem in the common block GEOTABLE will result.

CONSISTENCY CONDITIONS

When preparing the input file "geo.dat," the user is constrained by the law of conservation of mass. This law imposes a requirement that the functions $A_b(\delta)$ and $V(\delta)$ obey a certain consistency condition. This condition can be found by noting that the mass of the flare is equal to $\rho_s V(\delta)$. The time rate of change of this mass must be equal to $-\dot{m}$, since the burning of the flare is responsible for the decrease in flare mass. Substituting equation (1) for $\dot{m}$ results in

$$\frac{d}{dt}[\rho_s V(\delta)] = -\rho_s \beta A_b(\delta) \quad (5)$$

Because the solid flare density is constant, use of the chain rule results in

$$\rho_s \frac{dV(\delta)}{d\delta} \frac{d\delta}{dt} = -\rho_s \beta A_b(\delta) \quad (6)$$

Now the density can be canceled from each side. Equation (4) means that

$$\frac{d\delta}{dt} = \beta \quad (7)$$

Substituting equation (7) into equation (6) gives the result

$$\frac{d}{d\delta} V(\delta) = -A_b(\delta) \quad (8)$$

Equation (8) is a direct result of the law of conservation of mass. This equation means that the user must be careful when specifying the flare volume and burning surface area in the file "geo.dat," and must ensure that the numbers given there obey equation (8). Possible difficulties might arise if some approximations are used for the geometry calculations, and the approximations used for the volume and area are not consistent. Clearly, care should be taken in this area.

GROOVED SURFACES

At this point, the discussion of the modifications made to SMIRF is complete. But, the problem is not really solved, because the user is left with the problem of creating the file "geo.dat." To a certain extent, this is a problem for the user, since the algorithm for computing the contents of "geo.dat" will depend on the flare shape.

During the summer an approximate method of handling grooved surfaces was found. Since these grooves seem to be a common feature in flares, some space might profitably be devoted to explaining this approximation method.

Consider a surface such as the one shown in Figure 1. The flare surface contains a repeating series of grooves, which here are trapezoidal in cross section. The dimensions of the grooves are shown in the figure.

The effect of the grooves will be expressed in terms of two numbers. The first, the area augmentation ratio, α , is defined as the ratio of the area of the grooves to the area of the ungrooved surface. The mean surface location, shown as γ in Figure 1, is the position of the "average" surface, measured from the top surface. Together, these quantities will be useful for finding the areas and volumes of grooved shapes. The grooved surface can be replaced with the ungrooved "average" surface, the area and volume computed for this shape, and then the resulting surface area can be multiplied by α to find the burning area of the grooved surface.

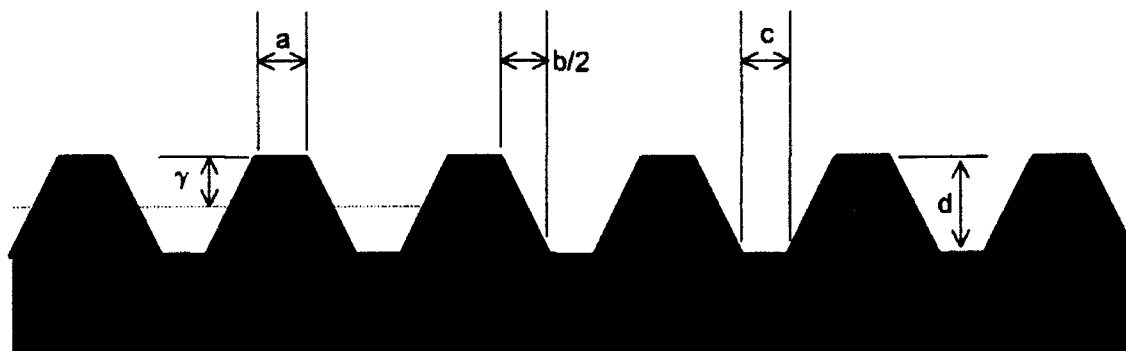


Figure 1. Consider a flare with a grooved surface. Here, the grooves repeat regularly and are trapezoidal in cross section. This figure defines the groove geometry used for the derivation in this section.

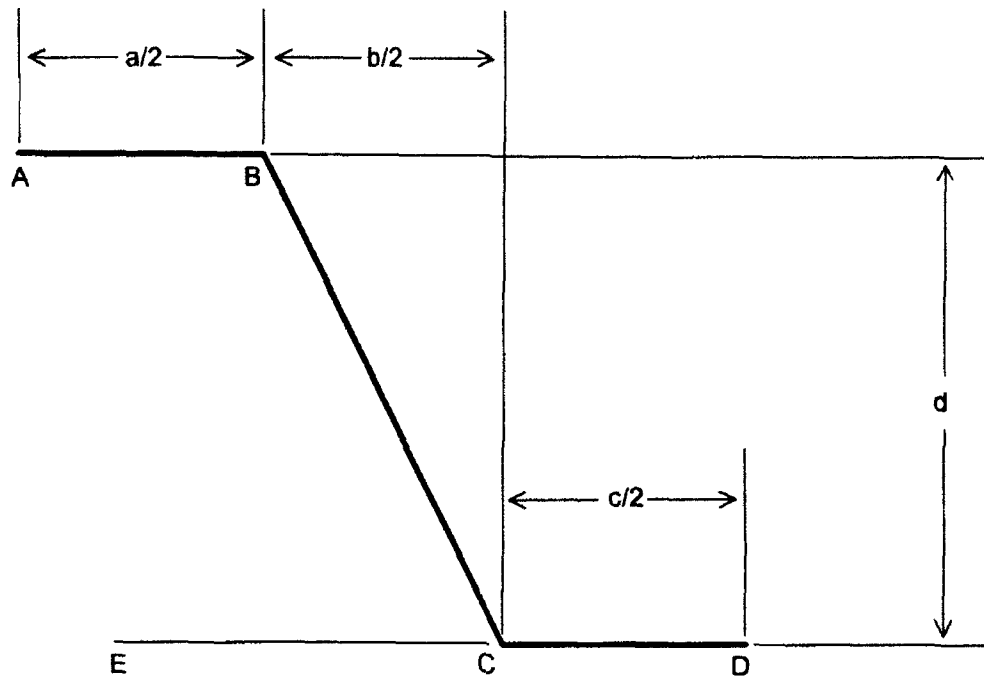


Figure 2. This is a close-up of the repeating unit of the groove geometry. The entire surface can be formed simply by reflecting this half groove about point D to form a whole groove, and then repeating that structure.

The analysis of the groove is simplified by examining only one half cycle of the repeating unit, as shown in Figure 2. The results of this analysis will be valid for the entire surface of Figure 1, based on mirror image symmetry and periodic repetition.

To get the desired result, the surfaces shown in Figure 2 must be allowed to propagate a distance δ into the solid, and then the area augmentation ratio and mean surface displacement must be found for the new surface position. Figure 3 shows the situation. The surface A B C D moves a distance δ to become the new surface A' B' C' D'. Actually, this picture is not totally true. Convex corners such as the one at B will remain sharp corners as the flare burns. But, concave corners such as the one at C will not. In reality, C' will be a curved surface rather than a sharp corner. However, this will not introduce serious inaccuracy into the results, and the approximation makes the analysis easier.

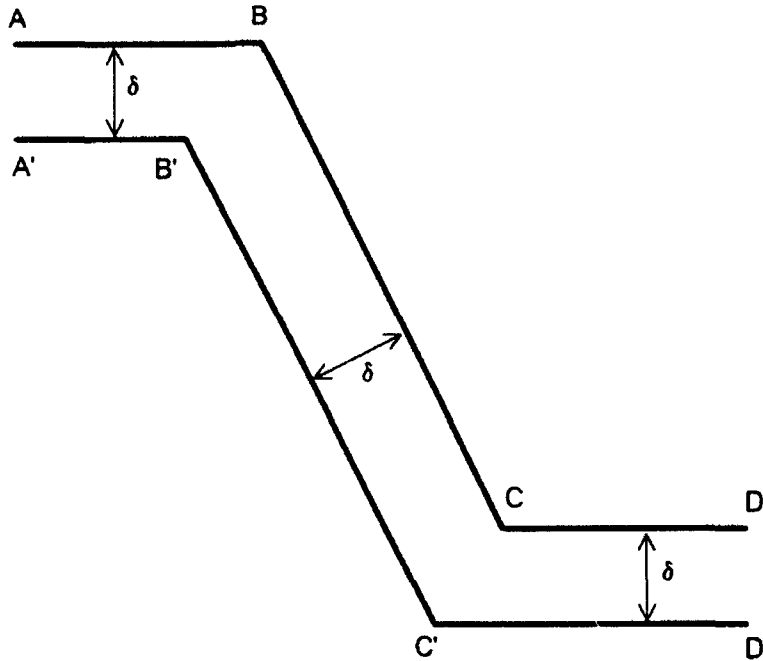


Figure 3. As the flare burns, the surface moves into the solid. Here, the unprimed points are on the original surface, and the primed points represent the surface after the flame has burned for a distance δ .

The position of the line $A' B'$ is found by requiring that it be parallel to line $A B$ and displaced by a distance δ . A similar requirement is imposed on the lines $B' C'$ and $C' D'$. This fixes the lines $A' B'$, $B' C'$, and $C' D'$. In turn, the intersections of these lines locate points A' , B' , C' , and D' .

Point A' will be located a distance δ directly beneath point A . The same will be true of points D' and D . The point C' will be located a distance δ below and a distance $\delta \tan(\theta/2)$ to the left of point C , where, referring back to Figure 2, θ is defined as the angle BCE . Point B' is located similarly with respect to point B . Thus, the new burning surface is (approximately) located.

From the location of the surface, expressions can be derived for the area augmentation ratio and the mean surface location. The area augmentation ratio is defined as the ratio of the true area of the surface to the area of its projection onto the horizontal axis. The result is:

$$\alpha = \frac{a + b/\cos\theta + c}{a + b + c} \quad (9)$$

The mean surface displacement is found from the requirement that the mean surface represents the surface level that would result if the surface were smoothed out, using the excess material removed from the high places to fill in the low places. Thus, the mean surface can be found by integrating the area underneath the line $A'B'C'D'$ and finding a constant surface level that has the same area underneath it. The result of this calculation is:

$$\gamma = \delta + d \frac{b/2 + c + 2\delta \tan(\theta/2)}{a + b + c} \quad (10)$$

These equations will not be valid forever. From Figure 3, it is clear that point B' is steadily moving to the left, while its mirror image on the other side of A' will be moving to the right. Eventually the two will meet, and the geometry represented by Figure 3 will no longer be valid. This will occur at a value of δ equal to

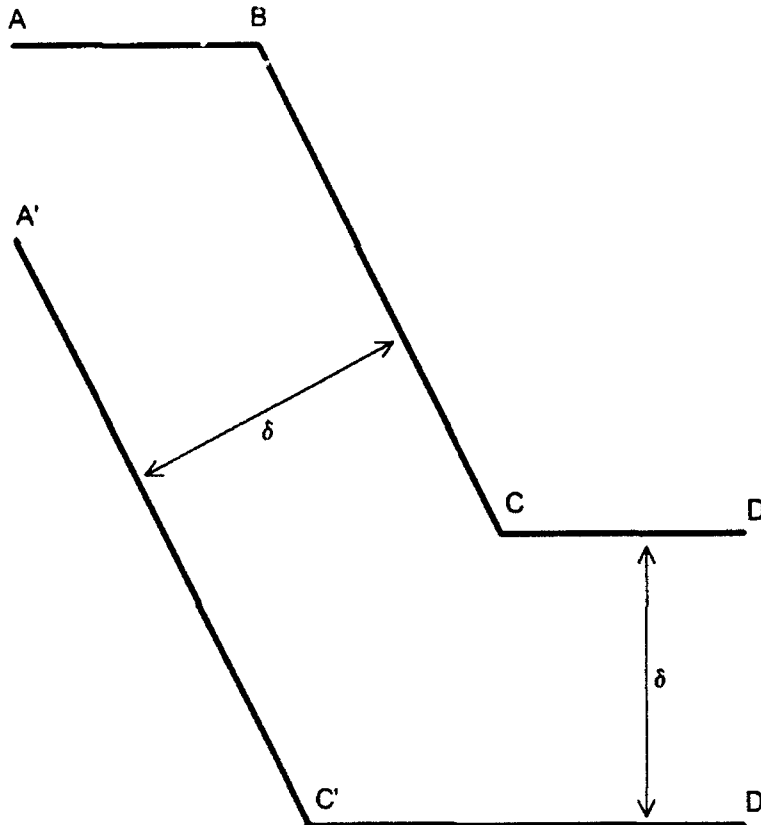


Figure 4. For $\delta > \delta^*$, the surface $A'B'$ has vanished and the surface looks like this.

$$\delta^* = \frac{a}{2 \tan(\theta/2)} \quad (11)$$

For all $\delta > \delta^*$, equations (9) and (10) are not valid, and the picture shown in Figure 3 is not right. When the surface A' B' totally vanishes, the surface will look like the drawing in Figure 4.

Under these circumstances, new expressions can be found that are analogous to equations (9) and (10):

$$\alpha = \left[\frac{a + b/\cos\theta + c}{a + b + c} - 1 \right] \left[\frac{\delta^{**} - \delta}{\delta^{**} - \delta^*} \right] + 1 \quad (12)$$

and

$$\gamma = \delta + d - \frac{(b/2 + (\delta - \delta^*)\tan(\theta/2))(d + (\delta - \delta^*)(1 - 1/\cos\theta))}{a + b + c} \quad (13)$$

In these equations, δ^{**} is the value of δ where point C', which is also moving steadily to the left, meets its mirror image and the grooves totally vanish. This criterion is given by

$$\delta^{**} = \delta^* + \frac{b}{2 \tan(\theta/2)} \quad (14)$$

When $\delta > \delta^{**}$, the grooves have completely vanished and the surface is smooth. In this case

$$\alpha = 1 \quad (15)$$

and

$$\gamma = \delta + d \quad (16)$$

These equations represent the solution. The quantities a , b , c , d , and θ are constants, fixed by the initial geometry of the grooves. The only real variable in these equations is δ . When $\delta < \delta^*$, equations (9) and (10) are used. When $\delta^* < \delta < \delta^{**}$, equations (12) and (13) are used. When $\delta > \delta^{**}$, equations (15) and (16) are used. The result is that α and γ are defined as functions of the variable δ .

EXAMPLE

As an application of the equations in the previous section, consider a flare which is roughly cylindrical in shape, but with grooves on the curved surface of the cylinder. Suppose that the initial length of the cylinder is L and the initial radius is R , where the radius is measured to the outer part of the grooved surface. For any given value of δ , the new length of the cylinder will be $(L - 2\delta)$ (both ends are burning) and the mean radius will be $(R - \gamma)$, where γ is defined in the previous section as a function of δ .

The aerodynamic surface area is defined as the total surface area of this cylinder, given by the sum of the areas of the cylindrical surface and the two ends:

$$A_A = 2\pi(R - \gamma)(L - 2\delta) + 2\pi(R - \gamma)^2 \quad (17)$$

The burning surface area is the same, except that the area of the curved cylindrical surface is multiplied by α , which represents the increase in surface area due to the grooves:

$$A_b = \alpha 2\pi(R - \gamma)(L - 2\delta) + 2\pi(R - \gamma)^2 \quad (18)$$

For this geometry, the volume and the aerodynamic volume are the same, and are given by

$$V = V_A = \pi(R - \gamma)^2(L - 2\delta) \quad (19)$$

In this way, the burning surface area, aerodynamic surface area, volume, and aerodynamic volume are found as functions of δ .

CONCLUSIONS

The SMIRF flare model was written to model the signatures of flares with simple geometrical shapes. This limitation has been removed, and the model will work for arbitrary shapes. To use this capability, the user of SMIRF has the responsibility of determining the variation of the flare surface area and volume as a function of the distance traveled by the flame front.

To keep the surface area and volume of internal passages from affecting the calculation of the drag coefficient, two additional variables were added to the model. These variables, the aerodynamic surface area and the aerodynamic volume, are intended to represent the gross shape of the flare and not include internal holes and the like.

Thus, the user has to provide a table which describes the variation of the burning surface area, the aerodynamic surface area, the volume, and the aerodynamic volume as functions of the distance traveled by the flame. SMIRF has been modified to read this data and use a table look up and interpolation scheme to find the geometrical quantities need to find the flare trajectory and signature.

An approximate method has been developed that allows the effect of grooved surfaces on the burning surface area to be found. The burning area of the grooved surface is calculated as the area of the surface without grooves times an area augmentation ratio.

REFERENCES

1. Denny, J. C. and Brown, D. G., "SMIRF, Signature Model for Infrared Flares, with Post Processor", April 19, 1992.
2. Norton, O. P., "The Effect of Carbon Particle Combustion on the Infrared Signature of a Magnesium-Fluorocarbon Flare", final report submitted to RDL for the 1991 AFOSR Summer Faculty Research Program, September, 1991.
3. Norton, O. P. and Bentley, H. T., III, "Combustion of Carbon Particles in the Plume of a Flare", presented at the 1991 meeting of the IRIS specialty group on infrared countermeasures. held May 14-16, 1991 at the Johns Hopkins Applied Physics Laboratory.
4. Dash, S. M., Pergament, H. S., Wolf, D. E., Sinha, N., Taylor, M. W., and Vaughn, M. E., Jr., "The JANNAF Standardized Plume Flowfield Code Version II, (SPF-II)", volumes I and II, technical report CR-RD-SS-90-4, U. S. Army Missile Command, July, 1990.
5. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., "Numerical Recipes; the Art of Scientific Computing", Cambridge, Cambridge University Press, 1986, pp. 89-92.

ACKNOWLEDGMENTS

This research was performed at Arnold Engineering Development Center (AEDC) under the 1992 Air Force Office of Scientific Research (AFOSR) Summer Faculty Research Program, administered by RDL, Inc. of Culver City, California. The author thanks the AFOSR for their sponsorship and acknowledges RDL for their capable administration of the program. The summer program at AEDC was coordinated by Captain Harold Martin, USAF. The research project was directed by H. T. Bentley, III of Sverdrup Technology/AEDC Group. Acknowledgment is also due to Janice Denny, Danny Brown, Martha Simmons, and Gary Ledbetter of Sverdrup for their help.

FOURIER TRANSFORM SPECTROMETRIC INFRARED
DETECTION OF AIRBORNE SOLVENTS

Randolph S. Peterson
Associate Professor
Physics Department

The University of the South
Sewanee, TN 37375

Final Report for:
Summer Research Program
Arnold Engineering Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

FOURIER TRANSFORM SPECTROMETRIC INFRARED
DETECTION OF AIRBORNE SOLVENTS

Randolph S. Peterson
Associate Professor
Physics Department
The University of the South

ABSTRACT

DOD, EPA, and DEA have quantitative and qualitative measurement needs for airborne solvents used in chemical processing and combustion facilities. These airborne solvents provide dim, extended sources for spectrometric studies that are not temporally stable by most laboratory standards. The use of a Michelson interferometer-based Fourier transform infrared spectrometer (FTIR) has been used to study a dilute gaseous solvent, acetone, in the laboratory under conditions expected in the field. Limitations of detection and opportunities for use of new technologies have been explored and comprise the bulk of this report.

FOURIER TRANSFORM SPECTROMETRIC INFRARED
DETECTION OF AIRBORNE SOLVENTS

Randolph S. Peterson

INTRODUCTION

Qualitative and quantitative field detection of airborne solvents is of interest to many agencies. For example, the Drug Enforcement Agency (DEA) is interested in the ability to detect airborne solvents used in illegal drug manufacture. A field-measuring instrument that can detect and unambiguously identify these solvents could provide the DEA with another legal investigative tool (to complement undercover detective work and informants) to find and shut down labs that illegally produce or refine controlled substances.

Detection of most solvents requires the measurement of absorption spectra in the 2 to 12 micron wavelength range¹. To identify the solvents requires an instrument with a resolution of at least 4 cm^{-1} over this spectral range². The dilute concentration expected of such solvents requires an instrument with large etendue³. The variable nature of field measurements necessitates the use of a measuring instrument that is capable of quick (at most a few seconds) measurement. The Michelson interferometer-based FTIR is a much better spectrometer under these conditions than the best slit-source diffraction grating instrument, even one with array detectors³.

Feasibility tests for swift detection and identification of airborne solvents have been performed on a Nicolet model 7199 FTIR. This instrument is a Michelson interferometer-based FTIR with a twelve years old computer system. Some limitations of this particular instrument are in general not limitations with newer FTIR models. However, the spectrometer itself is an excellent instrument, providing a good test instrument for future field studies of

airborne solvents.

FTIR BASICS

Light from a source is passed through a beam splitter, reflected from two front-surface mirrors and recombined, as shown schematically in Figure 1. The light waves interfere upon recombining. The interference is constructive if one beam travels a total distance that is equal to or is an integral number of wavelengths greater than the other beam. If the distances traveled by the two beams differs by any other distance, then some (or complete) destructive interference occurs. Varying this path difference in a continuous, systematic

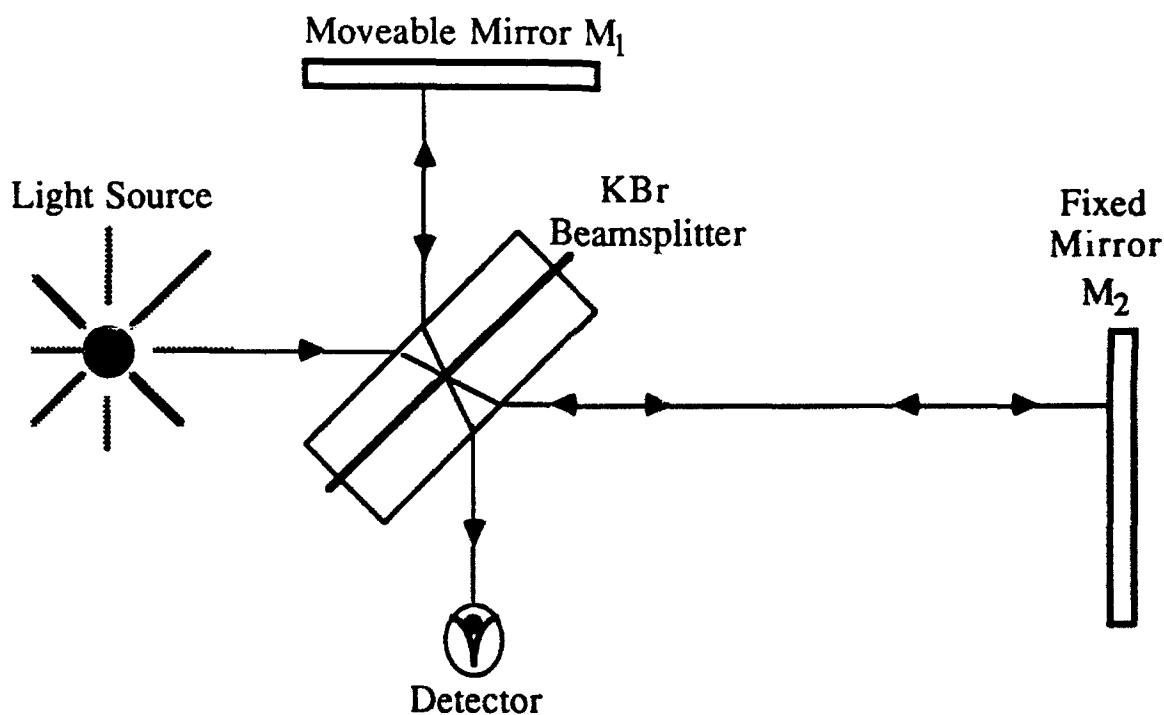


Figure 1. Schematic diagram of a Michelson interferometer

manner, by moving one mirror linearly at constant speed along the light beam path, produces alternating constructive and destructive interference patterns as a function of time, known as interferograms. An example of an interferogram is shown in Figure 2. The horizontal axis is the time axis, and represents a continuously increasing path difference for the two beams.

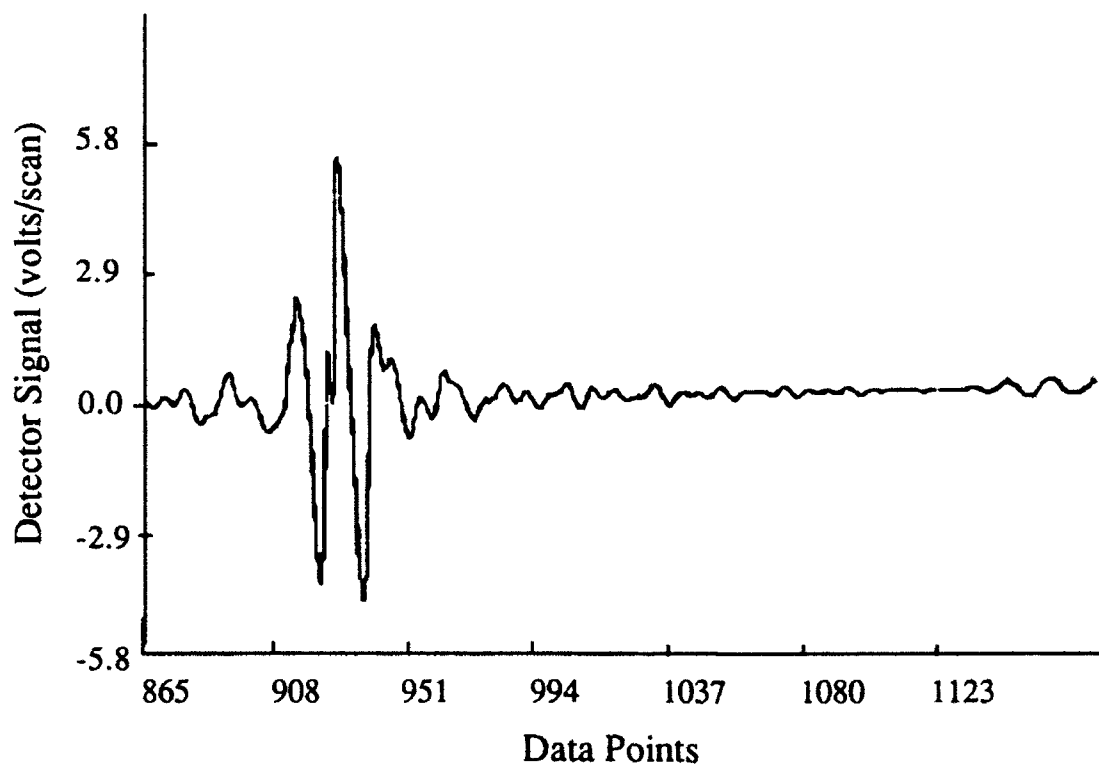


Figure 2. Interferogram from Nicolet FTIR spectrometer.

Traditionally⁴ the interferogram is digitized and Fourier processed to obtain the power spectrum³. A typical power spectrum is shown in Figure 3. The power spectrum displays the intensity (power) of each frequency emitted by a source and modified by emission or absorption in gases along the path of the

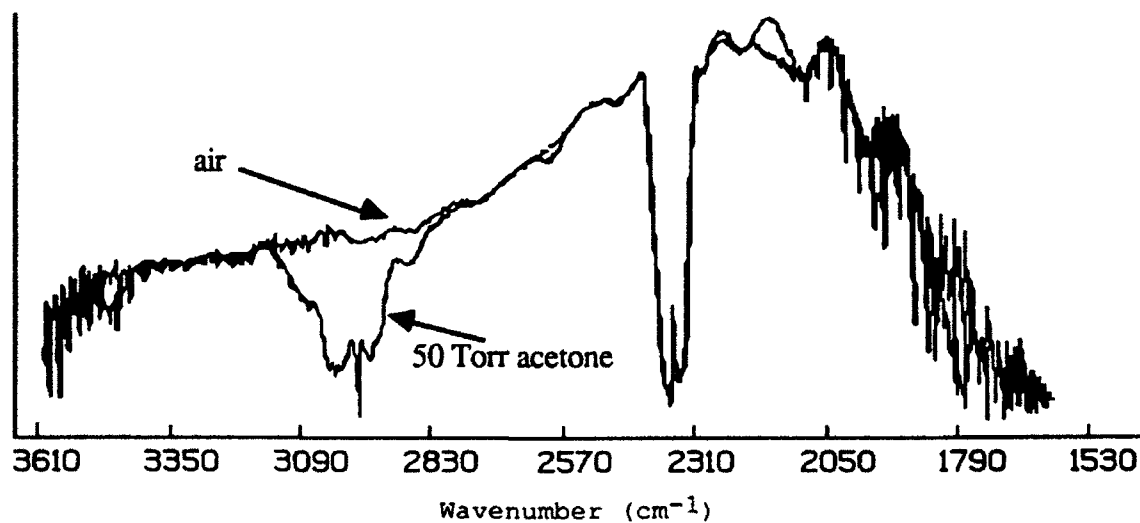


Figure 3. Power spectra for IR transmission through air and acetone.

infrared radiation from the source to the detector. By comparing the power spectrum of the observed sample to a power spectrum of the reference light, the absolute transmission through or absorption by the sample can be determined, as shown in Figure 4 for transmission through acetone.

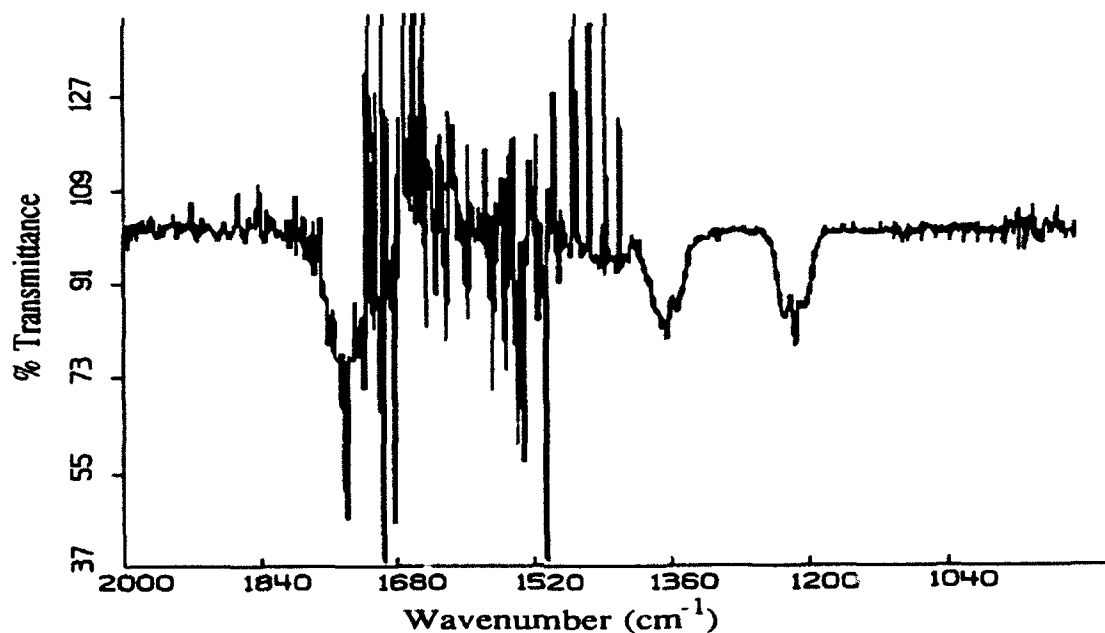


Figure 4. Transmission spectrum for acetone. The spectral resolution is 4 cm⁻¹.

Every type of molecule has a unique emission/absorption spectrum, allowing FTIR spectra to be compared to standard spectra for identification and quantification of the molecules present¹. From the measured percentage emission/absorption, the sample pressure multiplied by the observed path length can be determined from the Beer-Lambert law² (either in the linear or the exponential response regions).

Experimental Method

The FTIR spectrometer had not been operated in several years, so that the system required recalibration and some maintenance over the eight-week duration of this grant. The interferometer mirrors were aligned and a test of the instrument resolution was run on the transmission spectrum of carbon monoxide⁵. The CO spectrum is shown in Figure 5, with the two absorption

lines

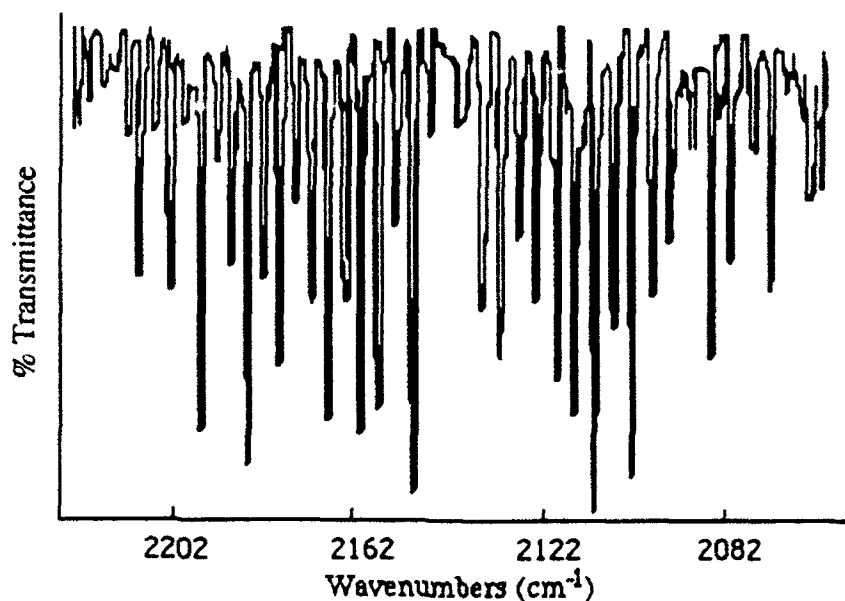


Figure 5. Absorption spectrum of CO in the infrared.

near the 0-0 point displayed in greater detail in Figure 6. The spectral resolution of the FTIR depends upon the number of interference maxima and minima observed, which in turn depends upon the total path traveled by the mirror. For the sampling rate used with this instrument, the resolution of the spectrum should be 0.25 cm^{-1} , which is the resolution observed in the spectrum of Figure 6. For the purposes of the research, a resolution of only

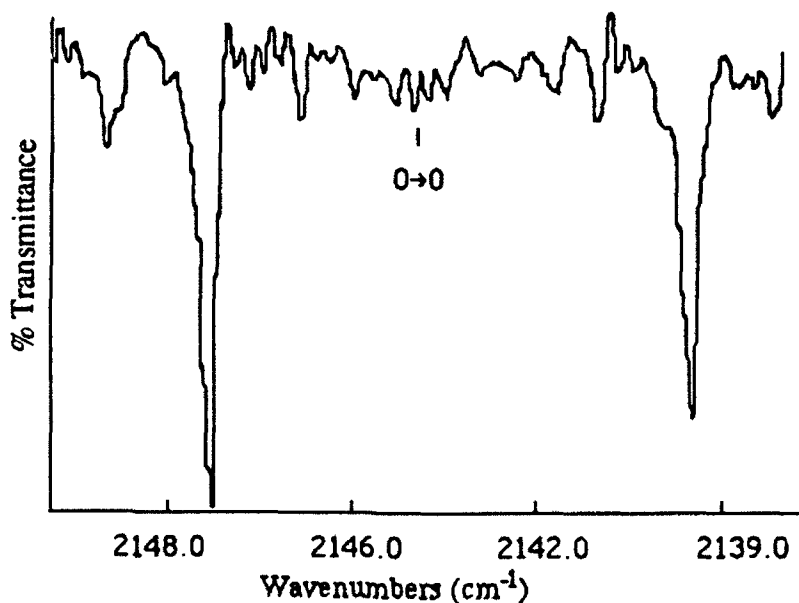


Figure 6. Details of Figure 4 spectrum. Instrument resolution is observed to be 0.25 cm^{-1} .

4 cm⁻¹ was needed, however, the higher-resolution measurement was a good check of proper mirror alignment.

Different sources of infrared radiation were used in the transmission measurements performed. In general infrared radiation was optically manipulated into a collimated beam of 5 cm diameter and transferred through the Michelson interferometer. The light from the interferometer was focused into a small diameter (about 1 mm) spot on the detector. A 10 cm long, cylindrical gas cell was introduced into the modulated beam of the FTIR between the interferometer and the detector.

The interference pattern was obtained from the amplified detector output with a 20-bit digitizer. The detector sampled the IR signal at a rate corresponding to a mirror movement of 632 nm, which was accurately measured by the interference pattern of visible light from a Helium-Neon laser. The interference pattern (interferogram) was stored in a computer and was Fourier transformed to obtain the power spectrum.

The pressure in the gas cell was monitored by a gauge accurate to a Torr. Pressure drifts during the acquisition of a spectrum were manually corrected since they were never more than a few Torr. The reference spectrum was the spectrum of a 0 Torr pressure gas cell. There was never any evidence from the power spectrum of gas impurities in either the gas spectra or the reference spectra.

A variety of IR sources, IR detectors, gas cell windows, and beam splitters were used with varying success during these experiments. Some equipment had degraded over the course of two years, and had to be replaced.

The final system used a controllable thermal emission source of varying , stable temperatures between 50 °C and 1,000 °C. The original beam splitter was replaced with a new KBr-Ge, allowing measurements at wavenumbers down to 700 cm^{-1} . Sapphire windows on the original gas cell were replaced by Irtran 2 windows for use below 2,000 cm^{-1} . The detector that was used for most of the final work was a Hg-Cd-Te detector, cooled by liquid nitrogen.

Results

The strongest absorption lines of acetone are in the frequency range between 1,100 cm^{-1} and 1,900 cm^{-1} , with weaker absorption lines near 3,000 cm^{-1} and 900 cm^{-1} .¹ The 3,000 cm^{-1} line was studied first, because the original optics absorbed the reference radiation in the other wavelength bands.

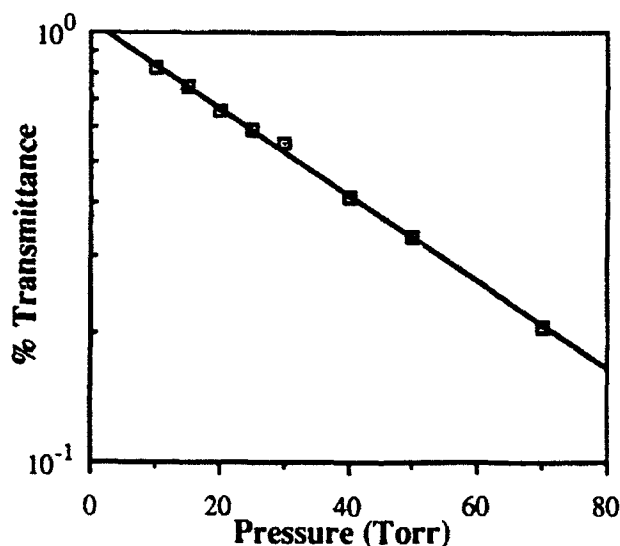


Figure 7. The transmission of 3,000 cm^{-1} IR as a function of the gas cell pressure.

The transmission of IR through a gas cell containing acetone was studied at 3,000 cm^{-1} as a function of gas pressure. The results are shown in Figure 7. The data fits an exponential reasonably well, as expected from the Beer-Lambert law, with an absorption coefficient of about 2/atm/cm. A sample

spectrum for these pressure variations is shown in Figure 8.

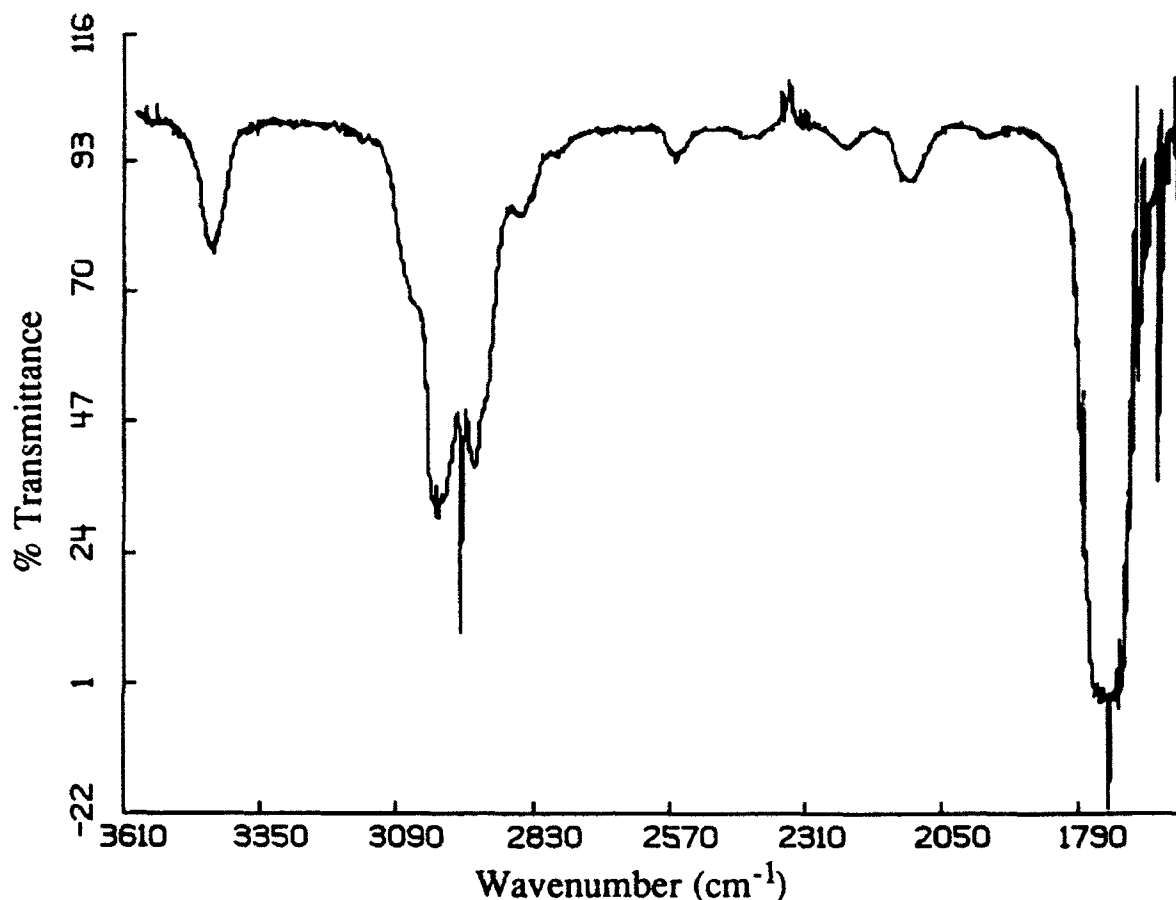


Figure 8. Transmission spectrum for IR passing through a 10 cm path of 100 Torr acetone gas. The spectral resolution is 4 cm^{-1} .

The 3,000 cm^{-1} line was easily distinguished for the lowest measured pressure of 10 Torr in a 10 cm long gas cell. The transmittance at this pressure was about 85%. If this is accepted as the minimum detectable amount, this represents about 3,000 ppbv for a 400 m long cloud of acetone. This is a reasonable value, but it represents the limitations of the pressure-monitoring equipment and not the limits of the instrument. This minimum could easily be less than 1,000 ppbv using the 3,000 cm^{-1} line, which is only a weak absorbing line in acetone. The stronger absorption bands were not studied as a function of pressure, but should easily provide minimum detectable limits less than 50 ppbv over a 400 m vapor region.

The strong absorption region (between 800 cm^{-1} and 1400 cm^{-1}) was

studied only briefly, since several new, nonabsorbing optical materials were needed for this frequency region that were not available with the FTIR spectrometer. These absorption lines were studied as a function of the temperature of a blackbody light source, which was the source for the IR reference beam. The results of a transmission spectrum of infrared radiation from an 800 °C blackbody passing through a gas cell of 100 Torr acetone gas pressure is shown in Figure 9.

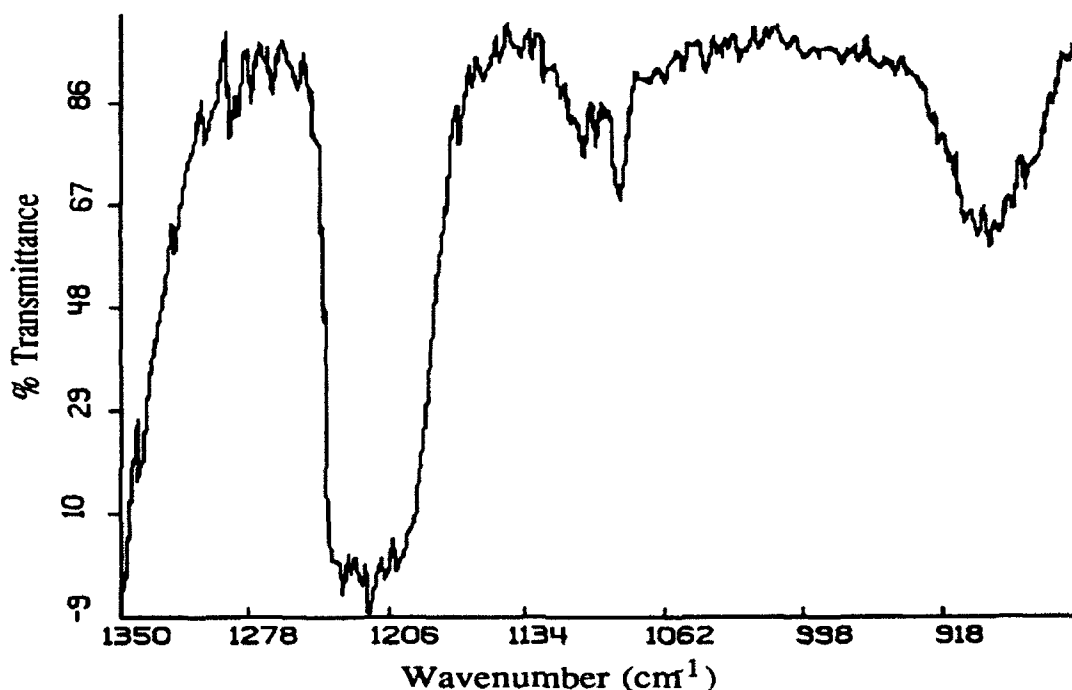


Figure 9. Absorption of IR from an 800 °C blackbody source by 100 Torr pressure of acetone in a 10 cm long gas cell. The spectral resolution is 4 cm^{-1} .

The absorption line near 1,200 cm^{-1} was easily observed (about 3% transmittance) using IR from a 100 °C blackbody radiation as the reference beam. Although reducing the blackbody temperature below 100 °C to near room temperature will significantly reduce the intensity of the reference beam, this absorption near 1,200 cm^{-1} should still be visible.

Because these FTIR measurements were performed in air, the strong absorption bands between 1,400 cm^{-1} and 1,800 cm^{-1} were never clearly resolved. Water vapor absorption lines are too strong and variable to

permit quantitative or qualitative measurements using the acetone bands in this frequency range.

The distinguishing absorption bands of acetone were studied for single scan FTIR spectra. Single scan spectra at 4 cm^{-1} resolution require only a few seconds for data acquisition. It was difficult to recognize any but the strongest absorption bands for 800 °C blackbody radiation and 50 Torr of acetone. Since only visual recognition of the acetone bands from the data was possible at the time, other techniques will improve the detection and recognition of acetone considerably.

Summary

Identification and quantitative detection of reasonable levels of airborne solvents (or pollutants) using an FTIR spectrometer requires enough resolution to be able to distinguish one organic molecule from another (4 cm^{-1} or less), data acquisition of seconds, a minimum detection concentration less than 100 ppbv for a 400 m path, and reference beams from blackbody (background) sources near room temperature.

The measurement of absorption of infrared radiation in acetone vapors has been studied as an example of the detection of an airborne solvent. In the time available for measurements, reasonable minimum detection limits (about 1,000 ppbv for a 400 m long cloud of acetone vapor at a frequency of $3,000\text{ cm}^{-1}$) were measured. These could easily be improved by about a factor of 50. Absorption bands of acetone were easily observed using infrared radiation from 100 °C blackbody emissions. The strong absorption bands of acetone were still visible from a single scan spectra. These single scans required only a few seconds to acquire at a resolution of 4 cm^{-1} .

References

1. Stadler Research Laboratories, *Stadler Standard Spectra, Gases and Vapors*, Philadelphia, PA, 1972.
2. Grant, Kagann, and McClenny, "Optical Remote Measurement of Toxic Gases," *J. Air Waste Manage. Assoc.* **42**: 18 (1992).
3. R. J. Bell, *Introductory Fourier Transform Spectroscopy*, Academic Press (New York), 1972.
4. R. T. Kroutil, J. T. Dittillo, and G. W. Small, "Signal Processing Techniques for Remote Chemical Sensing," in *Computer-Enhanced Analytic Spectroscopy*, Plenum Publishing Co., v. 2, 1990.
5. A. R. H. Cole, *Tables of Wavenumbers for the Calibration of Infrared Spectrometers*, 2nd edition, Pergamon Press, 1977.

**COMPUTATIONAL METHODS FOR CALCULATING
RADIATION TRANSPORT FROM PULSED X-RAY SOURCES**

**Richard M. Roberds
Professor
Engineering Science and Mechanics Department**

**The University of Tennessee Space Institute
Tullahoma, Tennessee 37388**

**Final Report for:
Summer Research Program
Arnold Engineering Development Center
Arnold Air Force Base, Tennessee**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

COMPUTATIONAL METHODS FOR CALCULATING RADIATION TRANSPORT FROM PULSED X-RAY SOURCES

Richard M. Roberds
The University of Tennessee Space Institute

ABSTRACT

Calculation of the transport of radiation from pulsed X-ray sources is a necessary part of radiation-effects testing. The characteristics of the radiation and the heterogeneity of the transport medium provide a challenge to traditional computational methods. This paper offers an introductory perspective of computational methods used for calculating photon transport associated with flash X-ray machines.

A brief review of the Monte Carlo (stochastic) approach is provided, but the focus of the paper is on the deterministic approaches of the discrete-ordinates and finite element methods. The derivation of the Boltzmann transport equation is reviewed along with simplifying assumptions (time-independent, multigroup) that are made. The discrete-ordinates method of discretizing angle is described, and mention is made of the numerical problems (ray effect and artificial dispersion) resulting from that approach. The finite element method as applied to radiation transport is briefly explained with emphasis on its promising aspects. Numerous references and a bibliography are provided which will enable the reader to gain an in-depth, comprehensive understanding of the subject if desired.

It is concluded that the Monte Carlo method is versatile and reliable and an essential backup method, but an inherently inefficient approach. Deterministic methods have specific advantages over Monte Carlo if they can be applied. The discrete-ordinates (S_N) method has been developed to a high level of sophistication, but the finite element method shows promise as a deterministic approach that may provide solutions to multi-dimensional problems which are impractical for the discrete-ordinates approach. While further development is required, the finite element method may eventually provide accurate deterministic solutions to problems with awkward geometries, and with an efficiency that surpasses the Monte Carlo method.

COMPUTATIONAL METHODS FOR CALCULATING RADIATION TRANSPORT FROM PULSED X-RAY SOURCES

Richard M. Roberds

INTRODUCTION

An essential part of an X-ray radiation effects test facility is the capability to accurately calculate the radiation field produced in and about the test article. The radiation field is characterized by such parameters as the time-dependent radiation flux (or time-independent radiation fluence) and radiation energy spectrum which give rise to radiation dose and dose rate. The parameters are determined by both the characteristics of the source of the radiation as well as the surrounding environment (geometry and materials of the medium) with which the radiation interacts on its way to the test article.

The term "radiation transport" is meant to describe the phenomenon of radiation propagation through, and interaction with, an arbitrary environment based upon a prescribed source of radiation. A single equation, the Boltzmann transport equation, precisely describes the phenomenon, and the task of calculating radiation intensities depends upon solving the transport equation. This equation, however, is an integro-differential equation with seven independent variables. Its exact analytical solution is not attainable, and one must resort to approximations to bring the equation into a tractable form for solution by numerical techniques. (An excellent reference on radiation transport and its associated computational methods is Duderstadt and Martin.)

The Boltzmann equation describes the transport of any of several physical forms of radiation. The radiation may be charged particles, neutrons, X-rays, gamma-rays, or low-energy photons (thermal or optical radiative transfer). Thus, its solution is of intense interest in many technical fields. Driven by such a variety of needs, computational methods for solving the transport equation have continued to develop over the years to high levels of sophistication.

While the computational techniques for solving the Boltzmann transport equation for the various radiation forms are mutually supportive to an extent, in general the specific problem to be solved dictates the approach that must be taken. For instance, this study is interested in X-ray and γ -ray (photon) transport. The transport of photons is very similar to neutron transport since in each case the equation is treating neutral (uncharged) particles. This similarity is fortunate since the nuclear community has supported a significant amount of research in neutron and γ -ray radiation transport over the years. However, the radiation transport understanding needed by flash X-ray facilities focuses on a particular photon energy range of interest that is at the low end of the energy range of interest for gamma ray transport. The specific geometries and transport media which comprise the X-ray test facility are also of interest. It is clear that the computational methods associated with flash X-ray facilities have their own unique approaches and developmental needs.

All radiation transport calculations have one need that is common, however -- the need to perform accurate calculations, in sometimes very complex geometries, without consuming excessive amounts of computer time or storage space. Even with the large super computers, running time and memory storage are most often the limiting factors in effective numerical computations.

Computational approaches for calculating the transport of radiation fall into two broad classifications: stochastic (Monte Carlo) methods and deterministic methods. In some instances the two methods may be combined to form a third "hybrid" category. Within these three areas there exists a large variety of techniques and approaches, many tailored to specific applications and problems.

Monte Carlo methods, if based on sound principles and techniques, may be relied upon to provide accurate results in any type of geometry or transport medium. In fact, Monte Carlo results are often used to represent the "correct" solution when comparing deterministic approaches. But Monte Carlo methods provide solutions only at pre-selected positions and require a substantial amount of computer running time to minimize statistical variance in the results. Consequently, in general, the Monte Carlo method is used when deterministic methods are inadequate.

The most widely used deterministic method for radiation transport analysis has been the discrete-ordinates S_N method and is applied using a finite-differenced Boltzmann transport equation. This method is the foundation of a series of radiation transport codes which began their development as one-dimensional codes in the 1950's. They were principally developed for neutron and γ -ray transport calculations and have reached a high level of sophistication. Current state-of-the-art, time-independent versions are the three-dimensional codes called TORT (RSIC CCC-543), developed by the Oak Ridge National Laboratory, and THREEETRAN (Lathrop, 1976) and THREEDANT which were developed by the Los Alamos Scientific Laboratory. Each of these 3D codes also has its one- and two- dimensional versions (DORT, TWOTRAN, TWODANT, respectively) (Lathrop and Brinkley).

Deterministic numerical methods provide transport solutions at each grid point established in the process of discretizing the spatial geometry. Solutions are not based upon a statistical determination and generally do not require the long running time needed by a stochastic approach, particularly from the perspective of CPU time per solution location.

Discrete-ordinates deterministic methods do have their limitations however. The most restrictive is the requirement that the problem geometry be rectangular, cylindrical, or spherical with boundaries placed along coordinate planes. Problems with irregular boundaries and irregular material distributions are difficult to solve accurately with the discrete-ordinates method. As an inherent consequence of the discretization of the directions along which radiation can travel, a "ray effect" is often created in multidimensional problems that produces spurious oscillations in the spatial distribution of the calculated flux density. Finally, the minimization of numerical truncation errors frequently calls for a fine angular and spatial mesh that causes the computational effort to become very large.

There is an emerging deterministic approach that is showing a great deal of promise as a computational method which encompasses the associated benefits of a deterministic solution, and it is not encumbered by the above-mentioned short-comings of the discrete-ordinates approach. This is the so-called finite element method (FEM). In this approach, the one-, two- or three-dimensional area of interest is divided into mesh cells or "finite elements." The shape and size of the elements may be tailored as necessary to meet the needs of the problem. While the sides of the elements are usually linear (or planes), a curvilinear boundary or complex geometry can be matched more accurately using finite elements than with a finite-differencing grid. Unlike the discrete-ordinates method, any arbitrary geometry can be treated using finite elements. Furthermore, FEM is free from the ray effect that plagues the discrete-ordinates technique.

The finite element method has been successfully applied to a variety of engineering applications including structural mechanics, heat transfer, fluid mechanics, and electrostatics. It has also been applied to radiation transport calculations (Beynon and Sehgal) of interest to nuclear-reactor physics, but it has not yet been developed to the general extent that it provides a practical and substantial improvement to radiation transport analysis. The problem of radiation transport associated with flash X-ray sources in a test-cell environment (complex 3-D geometry and media voids) presents a difficult challenge to discrete-ordinates methods, but it may be treatable by a finite element method. This would allow for accurate solution without the need to resort to Monte Carlo methods.

The purpose of this study was to review the state of the art of radiation transport computation as it may be applied to the calculation of X-ray transport from a pulsed radiation source test facility. The report is written so as to provide an entry-level introduction to the radiation transport Boltzmann equation and a basic understanding of the numerical approaches currently in use to solve the transport problem. The emphasis has been to review the deterministic computational methods of radiation transport and, in particular, the discrete-ordinates and finite element methods. For completeness, the Monte Carlo method is discussed briefly. A bibliography and references are provided to enable the reader to acquire a deeper knowledge of the subject areas as may be desired.

THE MONTE CARLO METHOD

The Monte Carlo method of calculating radiation transport is a stochastic method. Its approach is to employ statistical trials to determine the result of competing events in the lifetime of the radiation particle in question. By incorporating a large number of particle histories, each of which is random in nature, an estimate of some average particle behavior may be made. The accuracy of such an approach depends upon the extent to which the physics of the various processes a particle undergoes is understood, and the extent to which the cross sections (probabilities of occurrences) are known. It is important that the number of histories followed is large enough that statistical variance is minimized, that the biasing techniques employed are valid, and that computer round-off error is controlled. In principle, the

technique can be as accurate as observing experimental results since nature itself reacts according to probabilities. The Monte Carlo approach is often viewed as a computer-run, scientific experiment.

Monte Carlo is a powerful, reliable technique. Even when deterministic methods are unable to treat a problem due to complex geometry or unable to provide accurate results due to undesirable material properties, the Monte Carlo approach can be employed. The major drawbacks to Monte Carlo, however, are two-fold: 1.) it is inherently slow, requiring relatively large computer resources (running time and memory) to follow a statistically relevant number of particle histories through many events, and 2.) the answer is always a statistical estimate of the correct value rather than a precise value. Furthermore, results are obtained at preselected locations and, in comparison to deterministic approaches which provide a solution at each grid point, they are relatively inefficient when considering computer time and memory.

Many good references are available which explain the Monte Carlo method for radiation transport calculations. A primer on the subject is offered by Turner et al., and several textbooks dedicate a chapter to the subject if the reader is interested in a relatively succinct presentation (Chilton et al.), (Duderstadt and Martin), (Schaeffer). One of the best in-depth references is Spanier and Gelbard.

A widely used Monte Carlo computer code for time-integrated photon radiation transport calculations is the Integrated TIGER Series or ITS code (Halbleib and Mehlhorn), (RSIC CCC-467). This code may be operated in one, two, or three dimensions as the TIGER, CYLTRAN, or ACCEPT codes, respectively. Its approach is to couple electron and photon transport with or without the presence of macroscopic electric and magnetic fields of arbitrary spatial dependence. It may be used to calculate bremsstrahlung generation from an electron beam in the diode of a pulsed-power source, the transport and deposition of photons emanating from the high-Z anode (converter) of such a source, or both. It describes the production and transport of the electron/photon cascade from 1.0 GeV to 1.0 keV.

The Monte Carlo method follows individual interaction relationships between the particle in question and its environment, and the relative probability of the interactions occurring. It does not seek the solution of a mathematical equation. This is not true of the second class of computational methods, the deterministic methods. Deterministic methods begin with the equation that mathematically describes the transport phenomenon, applies simplifying approximations, and then numerically seeks the solution to that equation. Before proceeding to the deterministic class of methods, then, the transport equation will be presented along with an overview of the simplifying approximations that are applied.

THE BOLTZMANN TRANSPORT EQUATION

The Boltzmann transport equation is the fundamental equation that describes the transport of radiation from its sources and throughout the problem domain until it ends in escape or absorption. The essentials of its derivation and eventual simplification are presented

below. The reader is referred to several excellent texts which cover in further detail the elements of the following discussion (Bell and Glasstone), (Chilton et al.), (Goldstein), and (Schaeffer).

The equation is simply an expression of the conservation of a field of particles in motion flowing in and out of a volume V . If $N(\mathbf{r}, E, \Omega, t)$ is the number of particles per unit volume at position $\mathbf{r}$ with energy E and direction Ω , the objective is to solve for the number of particles per unit time that flow past the position $\mathbf{r}$. That is, if v is the speed of the particles, we are seeking the particle flux,

$$\Phi(\mathbf{r}, E, \Omega, t) = vN(\mathbf{r}, E, \Omega, t)$$

where, in SI units, Φ is expressed as $\frac{\text{particles}}{\text{m}^2 - \text{J} - \text{ster} - \text{sec}}$.

Now if:

- (a) is the net number of particles flowing out of a volume V
- (b) is the total number of particles suffering collisions in V per unit time (and are hence being lost due to their change in E and/or direction Ω)
- (c) is the number of secondary particles with energy E moving with direction Ω that are produced in V by all particle/medium interactions in V in a unit time with an incoming energy of E' and direction Ω'
- (d) is the number of particles that are being created per unit time with energy E and direction Ω by a source within V (call it $S(\mathbf{r}, E, \Omega, t)$),

then the change with time in the number of particles in the volume V is

$$\begin{aligned} \frac{\partial N}{\partial t} &= -(\text{loss}) + (\text{gain}) \\ &= -(a) - (b) + (c) + (d) \\ &= \frac{1}{v} \frac{\partial \Phi}{\partial t} \end{aligned}$$

In mathematical terms this expression of conservation is:

$$\begin{aligned} \frac{1}{v} \frac{\partial \Phi(\bar{\mathbf{r}}, E, \bar{\Omega}, t)}{\partial t} &= -\nabla \cdot \bar{\Omega} \Phi(\bar{\mathbf{r}}, E, \bar{\Omega}, t) - \sigma_t(\bar{\mathbf{r}}, E) \Phi(\bar{\mathbf{r}}, E, \bar{\Omega}, t) \\ &+ \int_0^\infty dE' \int_{4\pi} d\Omega' \sigma_s(\bar{\mathbf{r}}, E' - E, \bar{\Omega}' - \bar{\Omega}) \Phi(\bar{\mathbf{r}}, E', \bar{\Omega}', t) + S(\bar{\mathbf{r}}, E, \bar{\Omega}, t) \end{aligned} \quad (1)$$

Eq.(1) is the time-dependent Boltzmann transport equation where

σ_t is the total macroscopic cross section for a reaction per unit length

σ_s is the differential scattering cross section (the probability per unit length of a particle with energy E' and direction Ω' being scattered into energy E with direction Ω .)

In most instances when dealing with pulsed X-ray sources, we are principally interested in the time-integrated X-ray flux, or the "fluence". In this instance we acquire the fluence by integrating Eq.(1) over all time. Noting that

$$\int_0^{\infty} \frac{\partial N}{\partial t} dt = 0 \quad (\text{since the number of particles before the X-ray flash and well after the X-ray flash is zero.})$$

Therefore, defining

$$\int_0^{\infty} \Phi(\bar{r}, E, \bar{\Omega}, t) dt \equiv \Phi(\bar{r}, E, \bar{\Omega}) \quad \text{and} \quad \int_0^{\infty} S(\bar{r}, E, \bar{\Omega}, t) dt \equiv S(\bar{r}, E, \bar{\Omega}) ,$$

the time-integrated Boltzmann equation is obtained:

$$\begin{aligned} \nabla \cdot \bar{\Omega} \Phi(\bar{r}, E, \bar{\Omega}) + \sigma_t(\bar{r}, E) \Phi(\bar{r}, E, \bar{\Omega}) \\ = \int_0^{\infty} dE' \int_{4\pi} d\Omega' \sigma_s(\bar{r}, E' \rightarrow E, \bar{\Omega}' \rightarrow \bar{\Omega}) \Phi(\bar{r}, E', \bar{\Omega}') + S(\bar{r}, E, \bar{\Omega}) \end{aligned} \quad (2)$$

The first term in Eq. (2) is called the "leakage" or "streaming" term. It is usually written as $\bar{\Omega} \cdot \nabla \Phi$ which is accurate since ∇ operates only on the spatial dependence of Φ and not on Ω . The mathematical form of $\bar{\Omega} \cdot \nabla \Phi$ depends upon the spatial and angular coordinate systems chosen. Analytical forms of $\bar{\Omega} \cdot \nabla \Phi$ in the most commonly used orthogonal geometries are presented on pages 316, 317 of Chilton and pages 58,59 of Bell and Glasstone.

Photon Transport

In many photon-transport calculations the quantity of interest is the radiation intensity $I(\mathbf{r}, E, \Omega) = E \Phi(\mathbf{r}, E, \Omega)$. Radiation intensity is expressed in SI units as joules/m² (or, in time-dependent cases, watts/m²). In this case, the time-independent Boltzmann transport equation, Eq. (2), is modified by multiplication of E on both sides:

$$\begin{aligned} & \bar{\Omega} \cdot \nabla I(\bar{r}, E, \bar{\Omega}) + \sigma_t(\bar{r}, E) I(\bar{r}, E, \bar{\Omega}) \\ &= \int dE' \int d\bar{\Omega}' \frac{E}{E'} \sigma_s(\bar{r}, E' - E, \bar{\Omega}' - \bar{\Omega}) I(\bar{r}, E', \bar{\Omega}') + ES(\bar{r}, E, \bar{\Omega}) \end{aligned} \quad (3)$$

Multigroup Approximation

The steady-state transport equation of Eq. (2) or Eq. (3) is too complex to solve accurately even by a numerical approach. A chief problem is that the energy variability of the cross sections, which is considerable, is not analytically known. The usual approach is to render the energy dependence of the transport equation more tractable by an approach called the multigroup approximation.

The multigroup method begins by dividing the entire energy range into G contiguous energy intervals or "groups". The lowest energy of the highest energy group (Group 1) is E_1 , and as the group designator (g) increases, the energy of the energy groups decreases, e.g., $E_1 > E_2 > E_3$, etc. The transport equation is then integrated successively over each energy group to obtain a set of G energy-coupled equations, each equation treating the flux as if it had only a single energy or speed. The multigroup form of the transport equation is written as

$$\bar{\Omega} \cdot \nabla \Phi^g(\bar{r}, \bar{\Omega}) + \sigma_t^g \Phi^g(\bar{r}, \bar{\Omega}) = \sum_{s=1}^G \int_{4\pi} d\bar{\Omega}' \sigma^{s \rightarrow g}(\bar{r}, \bar{\Omega}' - \bar{\Omega}) \Phi^s(\bar{r}, \bar{\Omega}') + S^g(\bar{r}, \bar{\Omega}) \quad (4)$$

where

$$\Phi^g(\bar{r}, \bar{\Omega}) \equiv \int_{E_g}^{E_{g-1}} dE \Phi(\bar{r}, E, \bar{\Omega}) \quad (5)$$

$$S^g(\bar{r}, \bar{\Omega}) \equiv \int_{E_g}^{E_{g-1}} dE S(\bar{r}, E, \bar{\Omega}) \quad (6)$$

$$\sigma_t^g \equiv \frac{1}{\Phi^g(\bar{r}, \bar{\Omega})} \int_{E_g}^{E_{g-1}} dE \sigma_t(\bar{r}, E) \Phi(\bar{r}, E, \bar{\Omega}) \quad (7)$$

$$\sigma^{s \rightarrow g} \equiv \frac{1}{\Phi^s(\bar{r}, \bar{\Omega}')} \int_{E_g}^{E_{g-1}} dE \int_{E_s}^{E_{s-1}} dE' \sigma_s(\bar{r}, E' - E, \bar{\Omega}' - \bar{\Omega}) \Phi(\bar{r}, E', \bar{\Omega}') \quad (8)$$

To the extent the cross sections (group constants) can be accurately determined, the set of G multigroup equations is equivalent to the original energy-dependent transport equation. In the latter case, the G equations can be solved to acquire the G unknown $\Phi^g(\mathbf{r}, \Omega)$'s.

DISCRETE-ORDINATES METHOD

The method of discrete-ordinates (or discrete directions) is described in a host of references (Carlson, 1963), (Bell and Glasstone), (Chilton et al.), (Carlson and Lathrop), (Schaeffer). The method was first developed for radiative transfer calculations in stellar atmospheres (Wick), (Chandrasekhar) but was adapted for neutral-particle transport in nuclear engineering problems by B. G. Carlson in the mid 1950's. Carlson originally developed a discrete-ordinates method for nuclear reactivity calculations in 1955 (Carlson, 1955) which became known as the discrete S_N method and which has been successfully applied to a variety of transport problems. The method has received continual development over the years.

The essential basis of the discrete-ordinates approach is to discretize the angular distribution of the particle fluence by selecting specific directions along which the radiation is allowed to flow, and evaluating the fluence $\Phi(\mathbf{r}, \Omega)$ at these specific directions, Ω_j . If there are N such directions chosen, the single-group transport equation, Eq. (4), is transformed into N equations of the form

$$\bar{\Omega} \cdot \nabla \Phi^g(\bar{\mathbf{r}}, \bar{\Omega}_j) + \sigma_t(\mathbf{r}) \Phi^g(\bar{\mathbf{r}}, \bar{\Omega}_j) = \sum_{g'=1}^G 2\pi \sum_{i=1}^N w_i \sigma^{g'-g}(\bar{\mathbf{r}}, \bar{\Omega}_i - \bar{\Omega}_j) \Phi^{g'}(\bar{\mathbf{r}}, \bar{\Omega}_i) + S^g(\bar{\mathbf{r}}, \bar{\Omega}_j) \quad (9)$$

where the integral over all angles in the scatter-in term of Eq.(4) has been represented by a Gaussian quadrature:

$$\int_{4\pi} d\Omega' \sigma^{g'-g}(\bar{\mathbf{r}}, \bar{\Omega}' - \bar{\Omega}) \Phi^{g'}(\bar{\mathbf{r}}, \bar{\Omega}') \cong 2\pi \sum_{i=1}^N w_i \sigma^{g'-g}(\bar{\mathbf{r}}, \bar{\Omega}_i - \bar{\Omega}_j) \Phi^{g'}(\bar{\mathbf{r}}, \bar{\Omega}_i)$$

where w_i and Ω_i are known.

These N equations can be discretized in space by a finite-difference approximation technique. By considering enough directions and by using an appropriately fine spatial mesh, it is possible, in principle, to obtain a solution of the transport equation to any desired degree of accuracy.

Following the finite differencing of the spatial variance, the multigroup transport equation in three-dimensional rectangular coordinates with spatial cells of width Δx , Δy , and Δz would be (g notation suppressed for ease of notation) (Badruzzaman):

$$\frac{\mu_j}{\Delta x}(\Phi_R - \Phi_L) + \frac{\xi_j}{\Delta y}(\Phi_F - \Phi_N) + \frac{\eta_j}{\Delta z}(\Phi_T - \Phi_B) + \sigma_t \bar{\Phi} = \bar{S} \quad (10)$$

where Φ represents the average directional flux at a pre-chosen direction on the appropriate mesh cell surface (the subscripts R, L, F, N, T, and B represent the right, left, far, near, top, and bottom surfaces of the mesh cell, respectively). Following the convention of the TORT code (Rhoades and Childs), μ_j , ξ_j , and η_j are the direction cosines in the x, y, and z directions. $\bar{\Phi}$ and $\bar{S}$ are the cell-averaged fluence and total source ($S(r, \Omega_j) + \text{scatter-in term}$), respectively.

Auxiliary relations which relate the cell-averaged fluence as a weighted-average of the cell-surface fluences are then used to solve for all the unknowns. These auxiliary relations take on various forms and are used to provide the most accurate value of the cell-averaged flux (fluence) based upon considerations of the particular problem.

The TORT code uses the auxiliary relation known as the method of weighted difference if the R, Θ ,Z geometry is used. If the X,Y,Z geometry is chosen, there are three methods available: weighted difference, linear nodal, and linear characteristic (Rhoades and Childs). The linear nodal and linear characteristic methods are more accurate, but require greater memory requirements and longer CPU time since the number of unknowns is greater in the calculating scheme. In each case, the trade-off between using a higher-order flux solution method or simply using a finer grid mesh must be made. Further discussion of these methods is found in RSIC CCC-543, pages 32,33,37.

The problem to solve, then, can be seen to be a set of $G \times N \times (I \times J \times K)$ coupled simultaneous equations, where G is the number of energy groups, N the number of discrete directions, and I,J,K is the number of spatial mesh points in each dimension, respectively. The equations may be represented by a large matrix equation which is solved by inverting the matrix through a successive iteration scheme to reach convergence.

A Comparison: Discrete Ordinates Versus Monte Carlo

An interesting piece of work illustrates the time comparison between the discrete-ordinates and Monte Carlo methods. A one-dimensional discrete-ordinates code, ONETRAN, was used to perform coupled electron-photon transport calculations and were then compared to an equivalent Monte Carlo (TIGER code) calculation (Lorence et al.). Both a benchmark problem and a realistic shielding problem were treated with the two codes. The work forms an excellent comparison between Monte Carlo and discrete-ordinates methods in a one dimensional application since the same electron-photon coupling scheme was used in both codes. (The discrete-ordinates approach was a Sg/P7 approximation.) The energy deposited at mesh-point locations was calculated by both computational methods which agreed to within one-sigma standard deviation. The S_N solutions were not only accurate but were obtained with more than an order of magnitude less computation cost. (For the benchmark problem, the S_N calculation required 37 minutes of CPU time on a VAX 11/780; the Monte Carlo

required 20.5 hours. The shielding problem required one hour for the S_N and 30.5 hours for the TIGER code.)

Limitations of Discrete Ordinates

While the fundamental principles of the discrete-ordinates approach are straight forward, the user must be cognizant of certain limitations to apply the method effectively. One such awareness was discussed above in choosing mesh spacing and an appropriate nodal flux-solution method. Others are an awareness of two important problems which the discrete-ordinates approach may encounter in multidimensions--the ray effect and artificial dispersion.

As mentioned earlier, the ray effect may become dominant when computing fluence in or through large areas of low-density material (voids) such as vacuums, air, or gas atmospheres. Attempts to mollify this effect (Lathrop, 1968 & 1971), (Miller and Reed) have not been entirely successful, yet this is a very important class of problems for discrete-ordinates codes. Mesh sizes must typically be chosen to be large for this class of problems due to the large, sparse nature of the geometry, and this causes the problem to be susceptible to numerical approximation errors.

While ray effects are well known, if not well controlled, another phenomenon, artificial dispersion, is less known. This phenomenon is described in Appendix G of RSIC CCC-543. Dispersion occurs when a monodirectional source flows past a sharp-edged, opaque obstruction. A shadow should be cast on the lee side of the obstruction, but due to the numerical representation of the flux, a portion of the flux appears within the shadow. This is lateral dispersion and occurs whether there is scattering or not since it is a numerically induced phenomenon.

The user may minimize the effect of dispersion by choosing an appropriate method for calculating the cell-averaged flux. However, while the TORT users manual offers some suggestions for which nodal method to use when employing TORT, depending upon the problem to be solved, experience will likely be the best teacher.

These difficulties are mentioned to emphasize two points:

1. The development of an effective computational ability to apply discrete-ordinates methods to the radiation transport problems of radiation effects testing will require knowledge of radiation transport theory coupled with a mature level of experience using the discrete-ordinates codes.

2. Computational development by the testing facility should be an ongoing activity to increase understanding and improve computational methods associated with radiation transport calculations.

FINITE ELEMENT METHOD

Deterministic computational methods have several advantages over Monte Carlo methods, but they are not always suitable for use. The discrete-ordinates method represents a very high level of sophistication in deterministic methods, yet it is limited to geometries which

can be modeled by spheres, cylinders, or slabs. As mentioned in the preceding section, the method may also be hampered by areas of material voids which create dispersion effects and ray effects.

A deterministic method that has enjoyed a high-level of success in the fields of solid and fluid mechanics shows promise for also being able to treat radiation transport problems on systems of awkward, if not intractable, shapes, and it is not susceptible to the ray effect. This method is the finite element method (Chandrupatla and Belegundu).

The success of the finite element method (FEM) within structural mechanics led to its application to fluid mechanics and eventually to electrostatics and heat transfer. It was, however, fairly belatedly introduced to radiation transport calculations in nuclear reactor physics. Its reluctance to enter the radiation transport arena has been largely due to the dimensional problems involved. In radiation transport, in contrast to the earlier applications mentioned, there are three additional dimensions in the phase-space of the field variable--two for the particle direction and one for particle energy. This has presented an added complexity to its practical application to neutron or photon transport.

The finite element method began to receive serious interest by the nuclear engineering community in the 1970's, but it was not until the late eighties that the method reached a stage where it could undertake precise practical calculations on systems of irregular shape (Akroyd, 1987).

An excellent overview article, "The Why and How of Finite Elements" was written by R. T. Akroyd (Akroyd, 1981) in which he lists several attractive attributes for finite element methods. Of the seven he lists, five are significant to the application of this review:

1. The ability to treat awkward shapes
2. The ability to obtain approximate solutions which are optimal in some well-defined sense for the whole system
3. The freedom from the spurious solutions (ray effect) which can cause problems with discrete-ordinates methods
4. The ability to combine accuracy and versatility by achieving very high accuracy for a wide range of problems.
5. The ability to vary the degree of sophistication of a transport approximation used across a system according to the physics of the problem, e.g., the ability to interweave the simpler diffusion equation approximation in areas where it would provide accurate results.

The particular interest of this study is the computation of photon transport in complex geometries and in a medium which contains areas of minimal photon scattering, "void" areas which are conducive to dispersion and ray effects. The Monte Carlo method is currently the only practical approach that is available to treat such a problem system with any acceptable degree of accuracy. However, recent activity in the area of finite elements indicates a potential for developing FEM for use in such problems. For instance, developments of the finite element method have been concerned with irregular arrays and the minimization of the mismatch between the behavior of the continuum and its discrete model (Ackroyd, 1981); this is the problem issue that must be bridged by deterministic methods in treating difficult shapes.

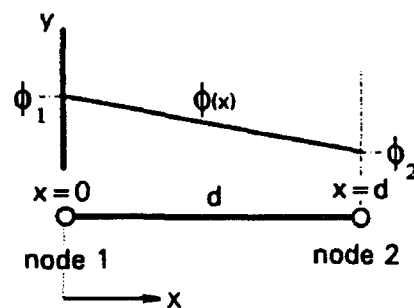
In a more recent practical development, a complete code based on FEM was used to study the attenuation of γ -rays in lead and water systems (Wood and de Oliveira). These developments support the notion that the finite element method would be fertile ground for exploring a deterministic approach to provide accurate calculations of photon transport in complex geometries.

Essentials of FEM

The essentials of the finite element method are these: The domain of the problem of interest is partitioned into a number of contiguous sub-regions (finite elements) by use of points (1D), lines (2D), or surfaces (3D). The surfaces may be curved if required, but they are specially tailored to meet the boundaries of the media and the physics of the problem. Figure 1 illustrates how the finite element mesh, using triangles to model a two-dimensional problem, can match the boundaries of a problem domain. Areas where large flux gradients are expected can be more finely meshed in a selective manner. Since the problem shown is rectangular, a finite-differencing scheme would also be able to match the boundaries, but a finer mesh is not able to be *selectively* placed. This results in less efficiency (greater storage and CPU running times). If there were curved geometries involved in a rectangular coordinate system, the finite-differenced mesh would not be able to match the boundaries, whereas a finer finite element triangle mesh would do much better.

The field variable (the unknown which is being sought) is described by a "local" approximation (an approximation within each element). In the case of multigroup radiation transport, the field variable is $\phi(\mathbf{r}, \Omega)$. This local approximation is almost always given by a polynomial which, in turn, is expressed in terms of the value of ϕ at the nodes:

$\phi(x) = \sum_i N_i(x) \phi_i$. By way of illustration, and for simplicity, assume only a spatial variation and a one-dimensional problem with a node at each end of the finite element (in this case a line),

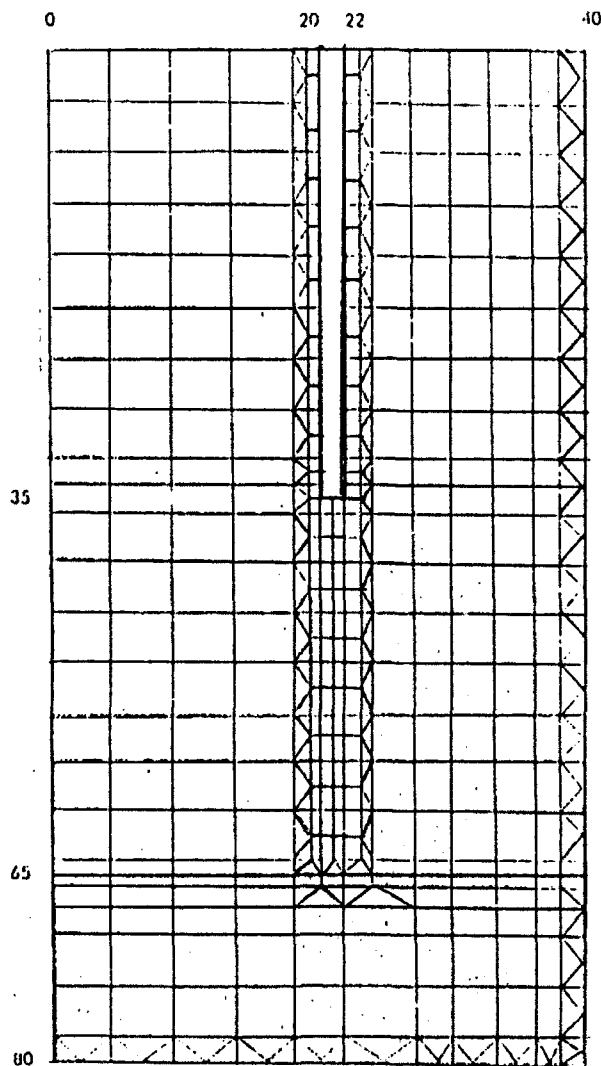
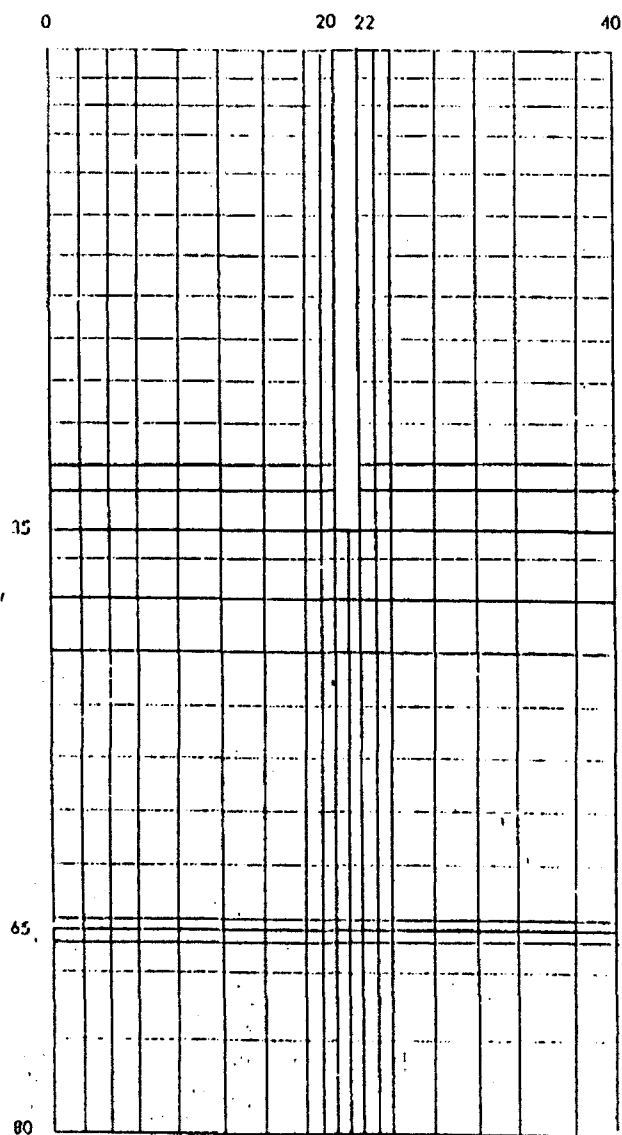
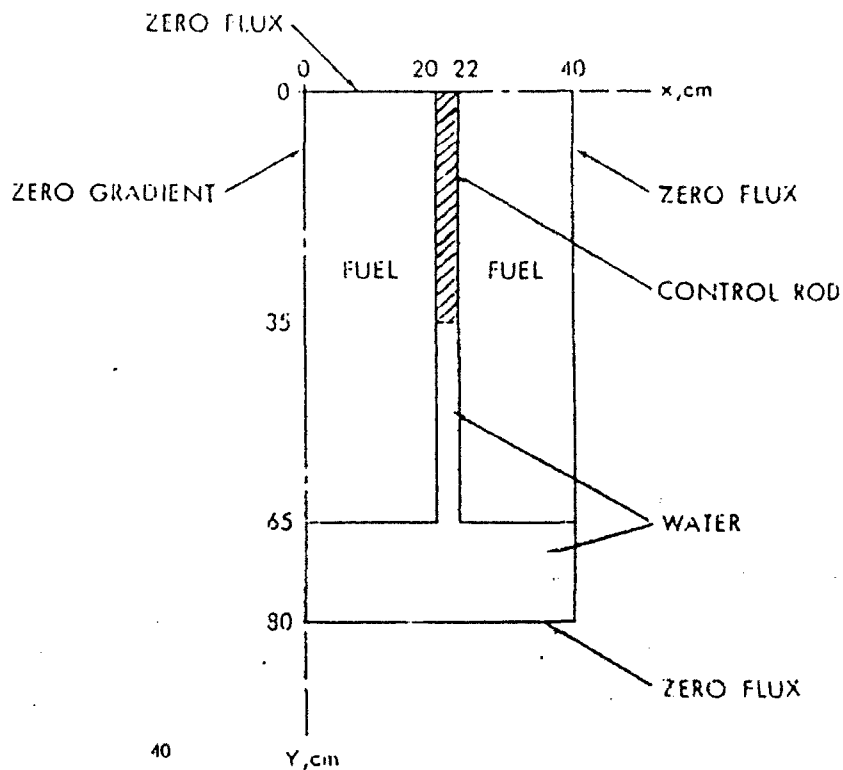


In particular, the field variable ϕ may be expressed as

$$\phi(x) = \left(1 - \frac{x}{d}\right) \phi_1 + \left(\frac{x}{d}\right) \phi_2$$

where ϕ_1 and ϕ_2 are the values of ϕ at node 1 and 2, respectively, and

Figure 1. A test problem (right) is finite-differenced in a rectangular-coordinate grid (lower left). This may be contrasted to a finite-element grid structure using triangles and rectangles (lower right). The latter can be seen to build a finer grid structure in a *selective manner*. (Numbers on the axes coordinate positions on the three figures.)



$(1 - \frac{x}{d})$ and $(\frac{x}{d})$ are the (linear) polynomials.

This approximation provides a linear change in ϕ from ϕ_1 at node 1 to ϕ_2 at node 2. However, a quadratic, cubic, or higher degree polynomial may be used.

A mismatch (residual) is defined for each element between the approximation ϕ and the exact solution Φ , and one of several possible mathematical means of minimizing this mismatch is used. For instance, if $R(\phi - \Phi)$ is an indication of the residual between the true value Φ and its approximation ϕ , the condition to minimize R for all the elements collectively, or to cause $R \rightarrow 0$, is imposed on the system. This minimizing of the residual results in a set of equations that may be simultaneously solved for the expansion coefficients of the polynomials (ϕ_1 and ϕ_2 in the above illustration).

The separate, specially tailored elements are then assembled into an interconnected whole as a model of the entire system. As the elements are joined, the continuity condition requires that the value of ϕ at each node of an element be the same as that for the adjoining element, since the node is the same physical point.

In summary, then, the general approach is that:

1. boundary conditions are placed on the global system (Φ 's defined at the boundaries),
2. an interconnection of each element's nodes is made by continuity conditions at the element boundaries,
3. the variance between element nodes within each element is described by polynomials that are optimized by the condition that $R(\Phi - \phi_0)$ is minimized, and
4. a system of simultaneous equations emerges that can be solved for the nodal values of ϕ .

These simultaneous equations are algebraic and may be expressed in the form of a large matrix equation. The numerical solution of this finite element equation in a computationally efficient manner often provides a considerable challenge. Fortunately, the finite element method usually results in sparse matrix systems which are often banded in such a way that computational and storage advantages can be realized (Martin et al.).

Mathematical Description

The finite element method is well grounded in mathematics. A thorough development of the mathematical foundation of a finite element treatment of the transport equation is provided by Wills. Simply stated, the problem is to solve an equation of the form

$$L\Phi = f \quad (11)$$

where L is an operator
 Φ is the dependent variable for which a solution is sought, and
 f is a function of the independent variables.

In the case of time-independent, multigroup radiation transport, Eq.(11) is the multigroup Boltzmann transport equation, Eq. (4), where:

$$L \equiv \bar{\Omega} \cdot \nabla + \sigma_1^s(\bar{r}, \bar{\Omega}) \circ ,$$

$$f \equiv \sum_{s'=1}^G \int_{4\pi} d\Omega' \sigma^{s'-s}(\bar{r}, \bar{\Omega}' - \bar{\Omega}) \Phi^{s'}(\bar{r}, \bar{\Omega}') + S^s(\bar{r}, \bar{\Omega}) \text{ and}$$

$$\Phi \equiv \Phi^s(\bar{r}, \bar{\Omega}) .$$

In mathematical terms, the desire is to project the exact solution of Eq. (11), including the boundary conditions, onto the approximating subspace of finite elements that describe the problem domain. This can be implemented using either a variational calculus approach (the Rayleigh-Ritz method) or by a more general approach, the method of weighted residuals (the most common form being Galerkin's method). It can be shown that these two methods are identical if the problem (operator) is self-adjoint (which is the case with the diffusion equation approximation of the transport equation). The so-called first-order transport equation (Eq. (2)) is not self-adjoint, and it must either be cast into a self-adjoint form (the "second-order" form of the transport equation), or the Galerkin method must be applied. Both approaches have been used (Martin et al.). For mathematical simplicity, the Galerkin approach will be outlined for a one-dimensional problem.

Weighted residual methods have been used for both the first- and second-order forms of the transport equation. The finite element approximation ϕ of Φ in Eq.(11) is found in the following way. $\Phi(x)$ is approximated by $\phi(x)$ within each element:

$$\Phi(x) \cong \phi(x) = \sum_{i=1}^m N_i(x) \phi_i \quad (12)$$

where $N_i(x)$ are the approximating (trial) functions, usually polynomials of an appropriate degree, and ϕ_i are the unknown expansion coefficients. The $N_i(x)$ may be viewed as basis functions that form $\phi(x)$.

Equation (11) states that $L\Phi - f = 0$. However, since $\phi \cong \Phi$, then we'd expect $L\phi - f$ not to be zero but to have a residual, R . That is, $L\phi - f = R$. The desire is to minimize R in a mathematically sound way, and through the minimization process, solve for the expansion coefficients. The method of weighted residuals defines a set of weight functions W_i in such a way that

$$\int_D [L\phi - f] W_i dD = 0 \quad i=1,2,\dots,m \quad (13)$$

where D is the problem domain.

The specific method of Galerkin is to choose a weighting function that is the same as the approximating (trial) function. So the Galerkin form of Eq. (13) is

$$\int_D [L\phi - f] N_i \, dD = 0 \quad i=1,2,\dots,m$$

This set of m equations can be solved for the unknown ϕ_i 's which can then be used to define $\phi(x)$.

The finite element approach is to perform this method for each finite element that comprises the global or problem domain. Hence, the spatial variable x is the local coordinate system of the element and not the spatial coordinate of the global system, which might be called X . The two coordinate systems must, of course, have a mathematical relationship when the elements are finally assembled to form the entire problem domain. This formalism results in a matrix equation which may be solved for the unknown ϕ_i 's.

CONCLUSIONS

As the state of the art of computational methods is viewed against the nature of the radiation transport problems encountered in pulsed-power X-ray test facilities, certain conclusions emerge.

1. The Monte Carlo method, while being less efficient than deterministic methods, is a versatile, reliable technique that can supply solutions in any well-defined problem scenario. It may also be used to supplement deterministic methods and as a comparison calculation. It is an indispensable tool for the computationalist interested in performing radiation transport calculations.

2. The discrete-ordinates method has been developed to a high level of sophistication and is the mainstay of deterministic methods. It has primarily been used for neutron and γ -ray transport and has not been used as widely in X-ray transport problems. It may experience difficulty in describing X-ray radiation transport in a test-cell environment or other problems associated with radiation effects testing. However, the advantages it has over Monte Carlo calculations and the maturity of its development require that it be available and applied where appropriate in performing X-ray transport calculations.

3. The finite element method has demonstrated promise for efficiently treating the radiation transport problems of radiation effects testing facilities. The advantages of the FEM approach is that it enables a deterministic treatment of the awkward geometries which are likely to be found in the problems faced by X-ray test facilities. Furthermore, it could provide higher accuracy by being able to more efficiently provide finer mesh spacing in critical areas, and it is not susceptible to the ray effect. These advantages lead to the conclusion that an effort should be continued to further develop the FEM for specific application to X-ray radiation transport.

BIBLIOGRAPHY

Ackroyd, R.T., "The Why and How of Finite Elements," *Annals of Nuclear Energy*, Vol. 8, pp. 539 to 566, 1981.

Ackroyd, R.T., Fletcher, J.K., Goddard, A.J.H., Issa, J., Williams, M.M.R., Wood, J., "Some Recent Developments in Finite Element Methods for Neutron Transport," *Advances in Nuclear Science and Technology*, V. 19, pp. 381-455, 1987.

Badruzzaman, A. and Fan, W.C., "Discrete Ordinates Methods for Radiation Transport", SAND-89-1909C, Sandia National Laboratories, 1989.

Bell, G.I. and Glasstone, S., *Nuclear Reactor Theory*, Van Nostrand Reinhold Co., NY, 1970

Beynon, T.D., Sehgal, B.R., "3rd International Seminar on Finite Element and Allied Methods in Radiation Transport: Shielding, Reactor Physics and Geophysics Applications," Special Issue on Finite Element and Allied Methods in Radiation Transport, *Progress in Nuclear Energy*, Vol. 25, No. 2/3, 1991

Carlson, B.G., "Solution of the Transport Equation by the S_n Method," USAEC Report LA-1891, Los Alamos Scientific Laboratory, 1955.

Carlson, B.G., "The Numerical Theory of Neutron Transport," *Methods in Computational Physics*, Alder, Fernbach, and Rotenberg, editors, Academic Press, New York, 1963.

Carlson, B.G. and Lathrop, K.D., "Transport Theory, The Method of Discrete Ordinates," LA-3251-MS, Los Alamos Scientific Laboratory, New Mexico, 1965.

Chandrasekhar, S., *Radiative Transfer*, Chapter II, Oxford University Press, 1950.

Chandrupatla, T.R., Belegundu, A.D., *Introduction to Finite Elements in Engineering*, Prentice Hall, Englewood Cliffs, New Jersey, 1991.

Chilton, A.B., Shultis, J.K., and Faw, R.E., *Principles of Radiation Shielding*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

Duderstadt, J.J. and Martin, W.R., *Transport Theory*, John Wiley & Sons, New York, 1979.

Goldstein, H., *Fundamental Aspects of Reactor Shielding*, Addison-Wesley Publishing Co., Inc., New York, 1971.

Halbleib, J.A., and Mehlhorn, T.A., "ITS: The Integrated TIGER Series of Coupled Electron/Photon Monte Carlo Codes," SAND 84-0573, Sandia National Laboratories, 1984

Lathrop, K. D., *Nuclear Science and Engineering*, Vol. 32, p. 357, 1968.

Lathrop, K. D., *Nuclear Science and Engineering*, Vol. 45, p. 255, 1971.

- Lathrop, K.D., "THREETRAN; A Program to Solve the Multigroup Discrete Ordinates Transport Equation in (x,y,z) Geometry," LA-6333-MS, Los Alamos Scientific Laboratory, Los Alamos, N. M., May 1976.
- Lathrop, K.D., and Brinkley, F.W., "TWOTRAN-II, An Interfaced, Exportable Version of the TWOTRAN Code for Two-Dimensional Transport," LA-4848-MS, July 1973.
- Lorence, L.J., Nelson, W.E., and Morel, J.E., "Coupled Electron-Photon Transport Calculations Using the Method of Discrete Ordinates," *IEEE Transactions on Nuclear Science*, Vol. NS-32, No. 6, December 1985.
- Martin, W.R., Yehnert, C.E., Lounce, L., and Duderstadt, J.J., "Phase-Space Finite Element Methods Applied to the First-Order Form of the Transport Equation," *Annals of Nuclear Energy*, Vol. 8, pp. 633-646, 1981.
- Miller, W.F., and Reed, W.H., *Nuclear Science and Engineering*, Vol. 62, 1977.
- Rhoades, W.A. and Childs, R.L., "The TORT Three-Dimensional Discrete Ordinates Neutron/Photon Transport Code," RSIC Computer Code Collection, CCC-543, January 1992.
- Roussin, R.W., "Structural, Heavy Coolant, and Material Cross-Section Data," *Progress in Nuclear Energy*, Vol. 13, No. 2/3, pp. 235-258, 1984
- RSIC Computer Code Collection, CCC-467, "Integrated TIGER Series of Coupled Electron/Photon Monte Carlo Code System, ITS Version 2.1", Oak Ridge National Laboratory, 1990.
- RSIC Computer Code Collection, CCC-543, "TORT-DORT, Two- and Three- Dimensional Discrete Ordinates Transport, Version 1.5.15," Oak Ridge National Laboratory, January 1992.
- Schaeffer, N.M., editor, *Reactor Shielding for Nuclear Engineers*, (TID-25951, National Technical Information Service, U.S. Department of Commerce), U. S. Atomic Energy Commission, 1973.
- Spanier, J., and Gelbard, E.M., *Monte Carlo Principles and Neutron Transport Problems*, Addison-Wesley Publishing Co., Inc., New York, 1969.
- Turner, J.E., Wright, H.A., and Hamm, R.N., "A Monte Carlo Primer for Health Physicists," *Health Physics*, Vol. 48, No. 6, pp. 717-733, June 1985.
- Wills, E.E., "A Finite Element Projection Method for the Solution of Particle Transport Problems with Anisotropic Scattering," Dissertation, University of New Mexico, Albuquerque, N.M., July 1984.
- Wick, G.C., "Uber Ebene Diffusionsprobleme," *Z. Physik*, Vol. 121, p. 702, 1943.
- Wood, J., and de Oliveira, C.R.E., "A Finite Element Study of Gamma Ray Transport," *Annals of Nuclear Energy*, Vol. 17, No. 4, pp. 195-205, 1990.

CIVIL ENGINEERING LABORATORY

CHARACTERIZATION OF SEAGRASS MEADOWS IN ST. ANDREW (CROOKED ISLAND)
SOUND, NORTHERN GULF OF MEXICO: PRELIMINARY FINDINGS

Sneed B. Collard
Professor
Biology Department

The University of West Florida
11000 University Parkway
Pensacola, FL 32514

Final Report for:
Summer Research Program
Headquarters Air Force Civil Engineering Support Agency

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

CHARACTERIZATION OF SEAGRASS MEADOWS IN ST. ANDREW (CROOKED ISLAND)
SOUND, NORTHERN GULF OF MEXICO: PRELIMINARY FINDINGS

Sneed B. Collard
Professor
Biology Department
The University of West Florida

Abstract

Seagrass meadows in St. Andrew (Crooked Island) Sound, a shallow marine lagoon bounded by Tyndall Air Force Base, Florida, were studied from mid-June to mid-August 1992. The study site was located at the narrowest part of the Sound about 3 km from its eastern origin. Seagrass meadows were monocultures of Thalassia testudinum and Halodule wrightii. Beds of the two species were often contiguous, with little or no overlap in distribution. Water temperatures during the study period ranged from 29.9 to 34.4 °C; salinity ranged from 30.1 to 36.0 ppt. At the deepest station (P5 CTL) water depth ranged from 40 to 118 cm. A pycnocline was often present at mid-depths in deeper water. Currents in the Sound are complex, and react instantaneously to wind stress. Seagrass densities ranged from 951 leaves m² at the shallow T. testudinum site (P3 CUT) to 2078 leaves m² at the deepest T. testudinum site (P5 CUT). Reliable measures of H. wrightii densities were not obtained. Growth rates ranged from -0.3 cm day⁻¹ to 2.5 cm day⁻¹. Fish and macroinvertebrate species diversity was high compared to other seagrass communities described in the literature. During a three day period a total of 57 species of fishes were collected using a small trawl and a crab scrape. This compares to a total of 128 species caught during both day and night trawl hauls over a one year period in the entire St. Andrew Bay estuarine complex. Heat and low tide exposure stress caused short-term diebacks in the shallowest beds, and contributed to a mass mortality of the sea urchin Lytechinus variegatus. Dominant macroepibenthic invertebrates included sea urchins, the Florida crowned conch, the horn shell, Bittium varium, blue crabs, hermit crabs, paleomonid shrimp, pink shrimp, mud crabs, hermit crabs, and arrow shrimp. Preliminary lists of species are included in the report. Infaunal species were inadequately sampled, and epiphytes were not sampled. Macroalgae were not found in or near the sampling area, an unusual, if not singular characteristic of the seagrass community in Crooked Island Sound (CIS). Large alligators are common in CIS both day and night, and pose some hazard to swimmers during nighttime sampling activities. Alligators may be important predators in CIS. Loggerhead and Kemp's ridley sea turtles were frequently observed from the entrance to Wild Goose Lagoon to the southeastern limit of the system. Sea turtles were presumably feeding on crabs and other crustaceans. Damage to seagrass meadows in the sampling area was caused by boat propellers, recreational scallop harvesting and intrusive sampling by the author. No direct or presumptive evidence of pollution from anthropogenic activities in the area was detected. Compared to other coastal ecosystems in the northern Gulf of Mexico CIS is relatively undisturbed, biologically diverse, and in many respects, appears to be ecologically unique. Results suggest that additional long-term studies of the system are warranted.

CHARACTERIZATION OF SEAGRASS MEADOWS IN ST. ANDREW (CROOKED ISLAND)
SOUND, NORTHERN GULF OF MEXICO: PRELIMINARY FINDINGS

Sneed B. Collard

INTRODUCTION

St. Andrew (Crooked Island) Sound¹ is a shallow marine lagoon located along a rapidly developing and ecologically sensitive portion of the Florida Panhandle. Compared to the St. Andrew Bay estuarine complex, Crooked Island Sound has received little investigative attention (e.g., McNulty et al., 1972; Saloman et al., 1982; Stout, 1984; Anon., 1988; Myers and Ewel, 1990; Turner, 1990). Little or no published information exists on the abundant seagrass meadows of CIS (Collard and D'Asaro, 1973; Humm, 1973; Den Hartog, 1977; Phillips and McRoy, 1980; Burch, 1981; Saloman et al., 1982; Phillips, 1982; Deegan et al., 1986; Iverson and Bittaker, 1986; Dawes, 1987; Durako et al., 1987; Fonseca et al., 1987; Virnstein, 1987; Zieman, 1987; Wolff et al., 1988; Fonseca, 1989; Lewis, 1989; Zieman and Zieman, 1989; Shaffer, 1992).

Seagrass communities are among the most productive of all marine ecosystems (Day et al., 1989; Larkum et al., 1989; Kusler and Kentula, 1989; Zieman and Zieman, 1989; Kennish, 1990), and their major ecological functions are relatively well known (Durako et al., 1987). All seagrass systems studied to date, however, are unique in terms of their interactive physicochemical and biological components; all are highly variable in space and time (e.g., Zimmerman and Livingston, 1976; Livingston, 1977; Robblee and Zieman, 1984); and at all but very long time scales, seagrass communities are unpredictable with respect to their ontogeny, architecture and metabolism (e.g., Greening and Livingston, 1982; Magoon et al., 1985 et seq.; Larkum et al., 1989; Mann and Lazier, 1991; Hackney et al., 1992). The ecological importance of seagrass communities and their individually unique characteristics argue strongly in favor of long-term research programs, particularly in the fundamental areas of community production and trophodynamics (e.g., Proceedings of a Workshop: The West Florida Shelf, in press). Information is needed on the ecological role(s) of all major taxonomic and functional groups of microbial, floral and faunal inquilines and associates of seagrass ecosystems.

¹ To emphasize the distinctness of St. Andrew Sound and nearby St. Andrew Bay, the former is referred to as Crooked Island Sound (CIS), its local epithet.

Short-term environmental research programs have been correctly criticized (e.g., Collard, 1989, 1991a). However, even limited sampling in space and time can be of value in providing rapid estimates of ecosystem health and quality. In addition, information acquired in baseline studies may be used to determine whether further, more intensive, longer/larger scale work is warranted (e.g., Abate, 1992). A distinction is made between one-time-only ("snapshot") sampling which is often of little value, and baseline research. Baseline studies are precursors, not substitutes for long-term environmental research programs. The latter are essential to increase our understanding of marine communities (e.g., Livingston, 1977; Collard, 1991b).

Objectives

The principal objective of the study was to conduct a coarse-grained warm-weather survey/characterization of SAV (submerged aquatic vegetation) in Crooked Island Sound. The study was undertaken to provide baseline information for use in the design of a long-term seagrass research program based in CIS. Concurrent with the collection of baseline information, a second objective was to evaluate the types and magnitudes of ecosystem-level environmental impacts attributable to local operations on the seagrasses of CIS.

Seagrasses

"Seagrasses" are salinity-tolerant, grass-like flowering plants which spend most or all of their life cycles submerged in estuarine and coastal marine waters from sub-polar latitudes to the tropics (Humm, 1953). Seagrass communities play important roles in the ecology of shallow coastal waters, and their widespread decline in Florida is of considerable agency and public concern. The decline of the commercial shrimp, Penaeus duorarum, in Florida has been ascribed to the dredging of seagrass beds (den Hartog, 1977). According to the Florida Department of Natural Resources (1987), "Seagrasses are a valuable part of Florida's marine environment but they are disappearing at an alarming rate. Dredge and fill projects and degraded water quality, as well as other activities, are responsible for their precipitous decline." Many fishes (e.g., speckled trout) and commercially important invertebrates (e.g., scallops) require SAV as nursery areas and/or sources of forage (e.g., Heck and Thoman, 1984; Zieman et al., 1984; Wooters, 1990)). Seagrass beds are important benthic habitats, and provide substrates for epiphytic flora and animal periphyton. In a study in Texas bays, more than 340 animals were reported to consume seagrasses (Fry and Parker, 1979). In estuarine ecosystems, submerged and emergent plant communities link terrestrial, intertidal and aquatic communities. For example, SAV is a primary source of energy in detrital food webs. Seagrasses are slow to decompose, which may result in a "time release mechanism" in the detrital cycle. The importance of algal epiphytes in seagrass food webs has been emphasized by numerous workers (e.g., Zieman, 1987; Wolfe et al., 1988). Kikuchi and Peresk (1977) identified

four sub-habitats associated with seagrasses: sediment fauna, rhizome and stem biota (amphipods, bivalves and polychaetes), leaf periphyton (microbiota, anemones, hydroids, ectoprocts, crustaceans, echinoderms, nematodes, polychaetes, and gastropods), and nekton which swim among the leaves (e.g., crustaceans and fishes).

Location of Crooked Island Sound and the Sampling Area

St. Andrew (Crooked Island) Sound is located in the northern Gulf of Mexico at approximately 30° N, 85° 30' W. Crooked Island Sound is bounded to the north by Tyndall Air Force Base (including Raffield Peninsula), and to the south by two tombolos known as Crooked "Island". Waters of CIS and St. Andrew Bay are not directly coupled. However, communication between the two systems via Gulf of Mexico shelf waters clearly occurs, with possibly significant, but undescribed geochemical and biological consequences.

Crooked Island Sound is about 14.5 km long, 0.2 to 2.0 km wide, and about 19,300 km² in area. From "Crooked Island Cut" to the eastern limit of the Sound the water is shallow (< 1 m on average), with narrow, meandering channels as deep as 3-4 m during mean tides. The location of sand shoals is predictable only over relatively short periods of time. At present, one navigable pass ("Crooked Island Cut") connects CIS to the Gulf of Mexico near the center of Crooked Island. East of the central pass Crooked "Island" is connected to the mainland. A narrow, shallow pass at the western end of CIS is marginally navigable by small boats according to the Air Force Marine Patrol (pers. comm.).

Crooked Island Sound was assumed to be physicochemically and biologically heterogenous, and the portion sampled could not justifiably be assumed to be representative of the Sound as a whole. It was considered important, therefore, to identify the sampling sites with some precision in order to facilitate relocation of the site.

Location of the Study Area

The study area was selected after exploration of the Sound by boat. The site selected was between Raffield Peninsula and Crooked Island: 29° 57' 51" N, 85° 28' 37" W (USGS Beacon Hill 7.5 minute quadrangle map, 1982). The study area extends from the border of a Spartina alterniflora bed (mid-intertidal zone) backed by a Juncus roemerianus marsh on Raffield Peninsula 55 m SSE into Crooked Island Sound. The shoreward position of the sampling area is 390 m SE (140°) of the Tyndall AFB (TAFB) Bench Mark on Raffield Peninsula; 540 m SW of a public boat ramp at end of Research Road; and 175 m S of the TAFB "Site 9700" perimeter road.

Location of Sampling Area Reference Markers

Permanent reference markers constructed of PVC pipe were driven about 1.5 m into the substrate along a straight line from the shore of Raffield Peninsula south into Crooked Island Sound. Reference Post 2 (P2 PST) was

placed at the boundary between a small Spartina alterniflora bed and an unvegetated sand flat. Post 3 (P3 PST) was placed 13 m south of P2 at the boundary between the sand flat and the edge of a Thalassia testudinum meadow. Post 4 (P4 PST) was placed 12 m south of P3 at a very sharp boundary between the T. testudinum meadow and a Halodule wrightii meadow. Post 5 (P5 PST) was placed 16 m south of P4, at the boundary between the H. wrightii meadow and a T. testudinum meadow, which extended some 18-22 m further south into CIS.

The ragged outer edge of the P5 turtle grass meadow was about half the distance between Raffield Peninsula and Crooked Island, at the narrowest part (ca. 110 m) of CIS. Water depths measured on 3 July 1992 during a spring flood tide were 75 cm at P3, 77 cm at P4, 98 cm at P5 and 258 cm at the outer edge of the T. testudinum meadow.

Location of Sampling Sites

Three 1.0 m² sampling sites ("CUT", "CLN" and "CTL") were placed in a "trident" pattern around each reference post. P3-P5 CUT stations were located 5 m east of reference posts P3-P5; CLN stations were located 5 m west of the reference posts, and P3-P5 CLN stations were located 5 m south of their respective reference posts. A three inch PVC pipe was driven into the substrate at the center of each square to mark the nine stations. Sampling site posts were fitted with PVC "T" connectors set to protrude about 20 cm above the substrate during inter-sampling periods. The use of "T" connectors allowed the insertion of either vertical or horizontal PVC pipes of varying lengths for ease of site recognition and for establishing transects away from the primary sites.

METHODS AND MATERIALS

In keeping with the short duration of the project (two months), and the exploratory-descriptive objectives of the research, basic sampling strategies and gear were employed. The kinds of information collected or measured were limited in scope. In spite of these limitations, data reported are accurate to within the levels of confidence indicated. Several hundred photographs of the seagrass communities in the sampling area were taken with a Minolta Maxxim 7000. Additional underwater photographs were taken with a Nikonos Model V. A video camera was used to document much of what was observed in CIS. Because of formatting requirements, these photodocuments cannot be used in the present report, but have been archived for follow-on work. Narrative descriptions of important observations recorded photographically (e.g., the appearance of seagrasses in the sampling area over time; the presence of penaeid shrimp, sea urchins and crabs on top of seagrass leaves at night) are impossible to construct without reference to source photographs. Sampling units were 1.0 m² and were constructed from PVC pipe.

Tides and Currents-

The tidal prism in CIS was estimated by measuring the depth of water with a meter stick at fixed reference poles (P3-P5 PST) during 13 sampling periods. This method was considered accurate to within ± 5 cm because of waves and boat wakes. An effort was made to sample the full range of tides as determined by frequent visual observations of CIS water levels in fringing Juncus marshes. Tide level assessments were made both day and night.

Current velocities were estimated by the drift rate of swimmers past reference posts while sampling in the water. The extent to which Thalassia blades deviated from their normally upright orientation was directly related to current flow, but reliable velocity estimates could not be made from these observations.

Temperature and Salinity-

Temperature was measured by immersing a calibrated thermometer in a flask of water collected at the surface or bottom. Because thermometers with different scales were used, accuracy was judged to be ± 0.5 °C. Soak times were approximately one minute. Salinity was measured with a calibrated refractometer considered accurate to within ± 0.3 ppt.

Seagrass Density Estimates-

Prior to estimating seagrass densities or initiating cropping experiments, blade heights were measured in situ (nearest cm) with a meter stick at sampling sites P3-P5 CLN and CTL. Forty blades selected at random from within each square meter site were measured from P3 and P5 (Thalassia) sites. At site P4 (Halodule) it was not possible to measure individual blade lengths without cutting them. Instead, 20 "clumps" of blades at P4 CLN and P4 CTL sites were grasped by hand and their "average" length was estimated. Murky water from suspended sediments further compromised the accuracy of measurements at P4. It was concluded that accurate leaf height and density measurements of both Thalassia and Halodule beds can be obtained only by cutting leaves at the substrate surface. These measurements are inaccurate, and are not included in the Results section.

Leaves were cut about 5 cm above the substrate surface with grass clippers on 2 July at Stations P3-P5 CUT. Clippings were held submerged under water until they were preserved in 10% formalin-seawater, washed in tap water and conserved in 70% ethyl alcohol.

Efficiencies of cut leaf recovery were estimated to be roughly 95% at P3 (T. testudinum), 75-80% at P4 (H. wrightii) and 90% at P5 (T. testudinum). The total number of recovered leaves plus the estimated number lost during field collection in each of the three samples was counted to provide a gross estimate of SAV density m^{-2} . Forty leaf lengths from each sample were measured to the nearest cm and counted, and modal leaf lengths were calculated.

Cropping-Growth Experiments-

Thalassia and Halodule blades within each m² sampling frame (P3-P5 CUT) were cut ca. 5 cm above the substrate surface with grass clippers on 2 July, and again on 8 July to remove those leaves that were missed during the initial cropping, and to ensure evenness of baseline blade lengths within the sampling frames. As a result of the second cropping, some leaves were cut twice, and this may have influenced their subsequent growth rates. On 11 July the heights of seagrass blades were measured at CUT sites P3-P5, and recorded. On 14 July leaves from P3-P5 CUT sites were removed at the substrate surface. These sites were checked (for QA) on 16 July, and on 1 August, 20 leaf heights from each site were measured to the nearest cm. It was noted that CUT sites, because they were effectively "holes" in the surrounding meadows, trapped a thick layer of drift Thalassia leaves.

On 14 July wire frames with 13/16th in. meshes were constructed to rest 3 cm off the substrate. A square hole 20 cm X 20 cm was cut in the center. These "cages" were installed at each (P3-P5) CTL site on 14 July. Seagrass leaves originating beneath the cut-out square were fed through the hole and cropped at the surface of the wire mesh (i.e., 3.0 cm above the substrate). Leaf heights were measured to the nearest cm with a meter stick on 1 August.

Sand-Mud Flat Infaunal Collections-

On 7 July a single 1.0 m² sampling square was placed on a sand-mud flat 50 m southwest of P2 and 10 m south of a Juncus roemerianus stand at low tide (depth of water at P3 = 16 cm). Air temperature was 32.2 C, water temperature was 34.0 C, and salinity at P3 was 33.8 ppt.

A garden shovel was used to remove approximately the upper 20 cm of sediments within the PVC frame. Each of 30 small substrate samples was placed in a 1.0 mm-0.4 mm sediment sieve series, and sorted in seawater at the collection site. Animals retained by the sieves were washed into 10% formalin-seawater at the site. Except for large, conspicuous species, specimens were not identified or enumerated. Ten randomly selected sand-mud flat sites within a radius of 50 m of the m² site were qualitatively sampled by extracting a small (ca. 20 cm deep, 0.2 X 0.2 m² area) sediment sample with a shovel. These samples were washed and preserved as above.

Seagrass Bed Sediment Depth and Infaunal Collections-

A post-hole digger was used to collect three 30 cm deep core samples from each (P3-P5) CTL site to determine major macroscopic infaunal species and characterize sediments at these locations. The animals collected were preserved in 10% formalin-seawater and conserved in 70% EtOH. Specimens were not identified to species or enumerated.

Try Net and Crab Scrape Collections-

Multiple five and ten minute tows were made on 30 July near P4-P5 CTL sites with a "try net". The trawl was fitted with a 1/8 inch mesh codend liner. Trawling was done at steerage speed from the bow of a pontoon boat (i.e., with the boat in reverse gear), and sampled a swath some 2 m wide (Koenig, pers. comm.). A majority of samples were collected in T. testudinum beds, but H. wrightii beds were also sampled. Tows were made in water depths of less than 20 cm to about 2.5 m. The sampling period lasted from 0930 to 1430.

During mid-morning to early afternoon on 30 July and 7 August ca. 30 collections were made with a one meter epibenthic crab scrape (Leber and Greening, 1986) to sample fishes and invertebrates in the vicinity of P4-P5 CTL sites. Towing was accomplished as described for the try net; again, more tows were made in Thalassia than in Halodule beds. On the morning of 11 August the crab scrape was towed by hand by M. Kuperberg and the author in the area of P4-P5, and ca. 500 m to either side of the P5 CTL post.

Juvenile lutjanid and serranid fishes were either frozen for otolith analysis, or preserved in 5% formalin-seawater for gut content analysis by Koenig. Two voucher specimens of each species were preserved in 10% formalin-seawater for use in the present study.

Night Observations/Dipnet Collections-

Numerous "sweeps" were made with a long-handled 1/8 inch mesh dipnet at sites P5 (Thalassia), and P4 (Halodule) on 8 July between 1800 and 2230 CDT. Collections were preserved in ca 10% formalin-seawater in the field. Animals were not identified to species or counted. A headlamp was used while walking in the seagrasses, and an electronic flash was used to photograph/document observations.

RESULTS AND DISCUSSION

The primary objectives of the study -- a warm-weather baseline survey/characterization of seagrass meadows in Crooked Island Sound, and an initial assessment of 9700 area operations on the seagrass ecosystem -- were only partially met for reasons mentioned earlier. Limited resources significantly impeded the work, and time constraints reduced its scope. A lack of critical taxonomic literature at the site precluded the identification of most infaunal animals.

Tides and Currents- The maximum tidal range measured at the deepest sampling station (P5 CTL) during July - 11 August 1992 was 78 cm ("mean" depth $\pm$ 39 cm). At the shallowest station (P3 PST) the range was 71 cm. This difference is attributed to sampling error, although cross-channel dynamic height differences probably occur during certain wind and tide conditions.

Winds were west-southwesterly at 5-10 kt during most of June, July and August except during local thunderstorms and weak frontal passages. Typical monsoon-like winds inhibited the outflow of water during low tides in the Gulf

of Mexico, and increased water levels during high tides. There was no correspondence between predicted NOAA (1992) tides and observed tidal heights in CIS. During the few sampling periods when winds were easterly during ebbing equatorial tides, strong seaward currents (est. > 1 kt) were encountered across the entire width of CIS at all depths. Seaward currents were generally restricted to deeper channel waters during ebbing tropic tides. Strong currents were not observed during flood tides regardless of wind direction or velocity during the sampling period. However, beginning on 10 August, a week-long period of severe storms dramatically influenced the sampling area. On 10-11 August, a persistent anticyclonic gyre was observed just seaward of the narrowest part of CIS (i.e., the sampling area). As winds approached (and likely exceeded) 50 kts field observations were discontinued. Because of its shallow depth, the waters of CIS probably respond instantaneously to wind stress, which generates complex and variable current patterns. Currents stronger than those observed probably occur in response to tides and wind forcing.

Temperature and Salinity-

During the sampling period water temperature in CIS ranged from 29.9-34.4 °C. A weak thermocline was often observed at approximately mid-depth in the region of the sampling area from P4 PST through P5 CTL. The strength and depth of the thermocline varied with water depth, the extent of wind mixing and tide conditions. For example, at 0930 on 22 July at P5 CTL, depth of the thermocline was ca. 30 cm below the surface; at 1700 there was no detectable thermocline at the station. The strength of the thermocline at P5 stations ranged from 0.1-0.6 °C. At P4 stations, thermocline temperature differences ranged from 0.1-0.4 °C, and at P3 stations the water column was mixed, with one exception. On 1 August at 1300 CDT, the temperature at P3 CTL was 33.0 °C at the surface and 32.0 °C at the bottom in water 56 cm deep.

In contrast to temperature, a halocline was often observed to be stronger at shallow than at deeper reference stations. When a halocline was present at one station, it was also observed at the others. While it is apparent that the water column in CIS is partially stratified, no "clean" explanations are justifiable at the present time. Precipitation was insignificant during the study period (mid-June through 11 August), and other sources of fresh water (e.g., from ponds or streams) were not observed.

Seagrass Density Estimates-

The range of leaf height measurements was large at all sites (least so at CUT), and the selection of blades for measurements was, in spite of attempts to collect leaves randomly, probably biased toward the largest and smallest blades. Leaf density (the number of leaves with intact tips) at each of the CUT stations was estimated to be: P3 T. testudinum: Total number/m² = [903 + (5% lost)] = 951. Remarks: Very few dead or damaged leaves; periphyton

coverage light. P4 H. wrightii: Total number/m² not determined (25% est. lost). Remarks: Counts of individual leaves proved to be impossible. Thousands of small snails on the leaves, and many of thousands of leaves. The impression is that H. wrightii produces more above-ground biomass than T. testudinum. P5 T. testudinum: Total number m² = [1870 + (10% lost)] = 2078. Remarks: Fewer dead or damaged leaves than on P3-CUT; periphyton light, about the same as P3. Much larger number of gastropods on this sample. Seagrass densities in CIS are high (see Bittaker and Iverson, 1976).

Sand-Mud Flat Infaunal Collections-

Square meter and randomly collected infaunal animals were not identified for reasons noted. Observations of larger animals in the field suggested dominance by callinassid shrimp, synaptid holothurians and polychaetes. Conspicuously absent were macroscopic infaunal gastropods, bivalves, and peracarid crustaceans, especially amphipods. Of those few species of polychaetes identified in the field, Arenicola cristata, Axiiothella mucosa, Diopatra cuprea (?), Pectinaria gouldii, Syllis sp., Onuphis sp., Glycera sp. and a terebellid were common. Virtually all animals were found in the upper 15 cm of sediment. It is emphasized that unvegetated substrates are considered effectively unsampled in this study. Superficial collections suggest that quantitative sampling is warranted.

Based on visual observation and the absence of hydrogen sulfide odor, the depth of the redox potential discontinuity layer was greater than 20 cm below the surface. At increasing distances (> 10 m) from Spartina and Juncus stands apparently living rhizomes (of Thalassia ?) were found about 10 cm beneath the substrate surface. A 5-20 cm thick layer of non-living "peat-like" material was found close to the edge of P3 PST, the border of a shallow Thalassia bed. Presumably, the "peat" deposits were remnants of former seagrass meadows. At all sand-mud flat sampling sites a very thin, diffuse layer of reddish-colored bacteria was observed.

Seagrass Bed Sediment Depth and Infaunal Collections-

A thin layer of reducing sediments occurred just beneath the surface in Thalassia beds; below this thin black layer sediments were oxidized. Coarse white quartz sand was found some 15 cm beneath near-surface, much finer sediments. Sediments in Halodule beds were uniformly silt-sized and reducing to about 30 cm, the maximum depth sampled. The root system of Halodule was much shallower than that of Thalassia.

Upon superficial examination, the major large infaunal invertebrates of Thalassia and Halodule appeared to be similar. Sediments and rhizomes were dominated by the same (?) three large species of non-tubicolous polychaetes. An ophiuroid and a nemertean not found in Halodule was abundant in Thalassia.

As noted for sand-mud flat infauna, the seagrass meadow infauna is considered to be unsampled in this study.

Try Net and Crab Scrape Collections-

Of those large invertebrates identifiable in the field the most numerous collected with the trawl and crab scrape were Melongena corona, Fasciolaria tulipa, Pagurus bonairensis, Palaemon floridanus, Hippolyte zostericola(?), Thor sp., Palaemonetes spp., Neopanope spp., Alpheus sp., Penaeus duorarum, Anachis sp.(?), Tozeuma carolinense, Callinectes sapidus, Melampus sp. and Lytechinus variegatus.

Fishes collected with try net and crab scrape are summarized in Table 1. Surface salinity on 30 July (36 ppt.) was the highest recorded during the sampling period.

TABLE 1
FISHES COLLECTED IN CROOKED ISLAND SOUND WITH TRY NET AND CRAB SCRAPE

<u>SPECIES</u>	<u>COMMON NAME</u>
1.* <u>Lutjanus griseus</u>	grey snapper (early juveniles) (c) C/F
2.* <u>Mycteroperca microlepis</u>	gag (1)
3. <u>Archosargus probatocephalus</u>	sheephead (2 adults)
4. <u>Paralichthys albiquetta</u>	Gulf flounder (c)
5.* <u>Lucania parva</u>	rainwater killifish (vc) Z/E
6. <u>Sphoeroides nephelus</u>	southern puffer (f)
7.* <u>Opsanus beta</u>	Gulf toadfish (vc)
8. <u>Diplectrum formosum</u>	sand perch (f)
9.* <u>Cynoscion nebulosus</u>	spotted seatrout (fc) C/F
10. <u>Ancylosetta quadricellata</u>	ocellated flounder (fc)
11. <u>Achirus lineatus</u>	lined sole (1)
12. <u>Paralichthys squamilentus</u>	broad flounder (r)
13. <u>Trinectes maculatus</u>	hogchoker (f)
14. <u>Symphurus plagiatus</u>	blackcheek tonguefish (fc)
15. <u>Bairdiella chrysoura</u>	silver perch (c) C/M
16. <u>Cynoscion arenarius</u>	sand seatrout (f)
17.* <u>Hippocampus zosterae</u>	dwarf seahorse (vc)
18.* <u>Hippocampus erectus</u>	lined seahorse (1) Z/E
19.* <u>Syngnathus springeri</u>	bull pipefish (c)
20. <u>Anarchopterus criniger</u>	fringed pipefish (r)
21.* <u>Syngnathus scovelli</u>	Gulf pipefish (c) Z/E
22.* <u>Syngnathus louisianae</u>	chain pipefish (c)
23.* <u>Syngnathus floridae</u>	dusky pipefish (fc)
24. <u>Myrophis punctatus</u>	speckled worm eel (1)
25. <u>Monacanthus hispidus</u>	planehead filefish (r)
26. <u>Aluterus scriptus</u>	scrawled filefish (r)
27. <u>Gobiosoma robustum</u>	code goby (c) Z/E
28. <u>Gobionellus hastatus</u>	sharptail goby (fc)
29. <u>Chasmodes saburrae</u>	Florida blenny (fc)
30. <u>Hypsoblennius hentzi</u>	feather blenny (fc)
31. <u>Ophioblennius atlanticus</u>	redlip blenny (f)
32. <u>Acanthostracion quadricornis</u>	scrawled cowfish (f)
33. <u>Chilomycterus schoepfi</u>	striped burrfish (f)
34. <u>Pomatomus saltatrix</u>	bluefish (c) (gill net)
35. <u>Mugil cephalus</u>	striped mullet (vc) OD
36. <u>Orthopristis chrysoptera</u>	pigfish (vc)
37. <u>Synodus foetans</u>	inshore lizardfish (fc)
38. <u>Sardinella anchovia</u>	Spanish sardine (r)
39. <u>Harengula jaguana</u>	scaled sardine (c)
40. <u>Lutjanus synagris</u>	Lane snapper (f) juveniles
41.* <u>Eucinostomus argenteus</u>	spotfin mojarra (f)
42. <u>Dasyatis americana</u>	southern stingray (vc) (dip net)
43. <u>Gymnura micrura</u>	smooth butterfly ray (f) (seen)
44.* <u>Menidia peninsulae</u>	tidewater silverside (vc) (dip net)
45.* <u>Lagodon rhomboides</u>	pinfish (vc) OD
46. <u>Mugil curema</u>	white mullet (c) OD (gill net)
47. <u>Sciaenops ocellatus</u>	red drum (f) (gill net)
48. <u>Scomberomerus maculatus</u>	Spanish mackerel (f) (gill net)
49. <u>Eucinostomus gula</u>	silver jenny (c)
50. <u>Dasyatis sabina</u>	Atlantic stingray (f?)
51. <u>Rachycentron canadum</u>	cobia (ling) (1) (hook and line)
52. <u>Micropogon undulatus</u>	Atlantic croaker (c?) (hook and line)
53. <u>Carcharhinus limbatus</u>	blacktip shark (3) (hook and line)
54. <u>Sphyrna barracuda</u>	great barracuda (r?) (juvenile)
55. <u>Gobiidae</u>	3 unidentified species (c)
56. <u>Sciaenidae</u>	1 unidentified species (r)
57. <u>Gobiesox</u> sp.	damaged-in aluminum can

[vc = very common; c = common; fc = fairly common; f = few; r = rare (1-3)]
[H = herbivore; OD = omnivore-detritivore; Z/E = zooplanktivore/epifaunal
invertebrates; C/M = carnivore/micro-macro invertebrates; C/F = carnivore/
fishes and macrocrustaceans (Gilmore, 1987, Table 4)] [* SAV-dependent]

Note: Ogren and Brusher (1977) identified 128 species of fishes from bi-weekly day and night trawl hauls at 12 stations in the St. Andrew Bay system during the period September 1972 through August 1973. Day-only collections at the study site in CIS were made on 30 July, 7 August, and 11 August using a 2 m trawl net and a 1 m crab scrape. On 11 August the trawl was towed through grass beds by hand. Gill net and hook and line records were personally observed by the author.

Night Observations/Dipnet Collections-

Space limitations preclude detailed discussion of day-night differences in the appearance of CIS seagrass meadows. As has been reported by others (e.g., Zieman and Zieman, 1989), many of the epiphytic grazers, herbivorous and carnivorous invertebrates (e.g., Melongena, Lytechinus, Penaeus, Astrea, gastropods, xanthid crabs, blue crabs) found in, on or near the sediment surface during the day were observed at or near the surface of seagrass leaves at night. No general statement can be made about this behavior, however, since very great differences were observed from night to night under similar conditions of moonlight, and between seagrass species. Far more pink shrimp and crabs were observed on the surface of Halodule than on Thalassia at night. The bay scallop, Argopecten irradians, appeared to be more active at night, as were some of the predatory fishes and their predators (raptors, alligators, Nerodia clarkii, etc.). The Atlantic silverside was only observed at night, while pinfish and pigfish, commonly seen during daylight hours, were not commonly observed at night.

Significant Observations-

The sampling location was considered unusual in two respects: (1) T. testudinum grew in slightly shallower water than H. wrightii, as noted in the description of the sampling area; (2) no macroalgae whatsoever were observed in the upper reaches of Crooked Island Sound. To my knowledge, the absence of either drift or attached macroalgae (e.g., Gracillaria, Caulerpa spp.) in an ecologically functional subtropical seagrass meadow has not been previously reported. Drift Syringodium filiforme was collected, but manatee grass was not observed in the vicinity of the sampling site. A single Halophila was observed growing at P5 CLN. Growth and turnover rates of seagrasses at sites P3-P5 increased as temperatures increased. Heat stress and extended periods of exposure at low tides caused short-term, patchy diebacks of shallow seagrass beds and episodic mass mortalities of (Lytechinus variegatus). Recreational scalloping and boating activities, as well as the author's frequent, intrusive sampling in and near the sampling area caused at least short-term damage to the meadow as a whole, as documented in time-series photographs. Damage was less pronounced in denser, deeper Thalassia at site P5 CTL than at other

sites. The number of propeller scars increased from 1 July through 11 August. Seagrasses were destroyed by propellers, and no evidence of re-growth was observed. Hourly, diel, day-night, weekly and longer-term significant changes in the appearance and community activity within the seagrass meadows was observed. Changes at these time scales can only be appreciated when sampling intervals are very short, and long, uninterrupted periods of time are spent in the meadows.

Adult loggerhead and Kemp's ridley sea turtles were frequently seen near the sampling area (Collard, in prep.). Sub-adult and very large adult alligators were seen frequently near the sampling area and, contrary to published reports, should be considered important predators in the CIS seagrass ecosystem. Also contrary to published reports, the Gulf salt marsh snake, Nerodia clarkii, is active both day and night in marine waters of the Sound. On 7 August, two Tursiops truncatus were observed in the area of P5 CTL. In the area studied, meadows were dominated by single species, and ecotones between adjacent beds were unusually abrupt, possibly indicating a sharp change in physicochemical conditions not measured or observed during the study period. Episodic "invasions" of the ctenophores Mnemiopsis, Beroe and Bolinopsis, unquestionably influenced zooplankton abundance when they were present in large numbers. No evidence of gross pollution was seen in CIS from Air Force activities in the area. Fish and macroinvertebrate species diversity was high; seagrass densities were high; epiphytic cover was moderate; and tidal flushing all suggest that seagrass communities in CIS are exceptionally healthy, but at risk because of overuse and misuse by human recreational activities (e.g., jet skis, shallow draft boats, scallop harvesting).

ACKNOWLEDGEMENTS

Jimmy C. Cornette made the project possible, and his unconditional support and enthusiasm for the work are greatly appreciated. Jimmy is one of those increasingly rare individuals who appreciates and understands both alligators and molecules. My thanks to Air Force Academy Cadet Darin Danielson, and especially to Philip Dorsch for their assistance in the field. Suzanne, Tyler and Dennis Collard assisted in the field and laboratory at night, and on weekends and holidays. Chris Koenig and Mike Kuperberg contributed significantly to the study by furnishing gear and organizing trawl and crab scrape collections. Mrs. Hazel J. Hill (AFCEA/RAV) saw to my survival at remote site 9700, and has my thanks. Don Volnoff (USAF) provided logistical support, and managed to procure many of the items that I requested. Jim Spain (AFCEA/RAV) kindly loaned me microscopes and other laboratory equipment. Dianne Bateman (USFWS) loaned me a refractometer and formaldehyde; Carol Monti (Panama City WWTP) contributed ethanol and formaldehyde; Rosalie N. Shaffer (NMFS Library Specialist, Panama City) gave me open access to reference materials from her extensive data base. I thank the staff of RDL for

their attention to details. My thanks to Colonel Neil Lamb and the Air Force Office of Scientific Research for providing me an opportunity to conduct research at the Air Force Civil Engineering Support Agency.

LITERATURE CITED

Because of limits imposed on the maximum length of the report by the publisher it was not possible to provide an adequate review of the literature. Much of the literature is cited by the authors listed below. Complete source information can be obtained from the author on request.

Abate, A. 1992. Environmental rapid-assessment programs have appeal and critics. *Bioscience* 42(7):486-489.

Baker, J. M. and W.J. Wolff (eds.). 1987. *Biological surveys of estuaries and coasts*. Cambridge Univ. Press, 449 p.

Capuzzo, J.M. and D.R. Kester (eds.). 1987. *Oceanic processes in marine pollution*. Vol. 1. *Biological processes and wastes in the ocean*. R.E. Krieger Publ. Co., Malabar, Florida, 265 p.

Carleton, J.T. 1989. Man's role in changing the face of the ocean: Biological invasions and implications for conservation of near-shore environments. *Conserv. Biol.* 3(3):265-273.

Collard, S.B. 1991a. The Pensacola Bay System: Biological trends and current status. Northwest Florida Water Manage. District, Surface Water Improvement Manage. Prog. Water Resour. Spec. Rep. 91-3, 181 p.

Collard, S.B. 1991b. Management options for the Pensacola Bay System: The potential value of seagrass transplanting and oyster bed refurbishment programs. Northwest Florida Water Manage. District, Surface Water Improvement Prog. Water Resour. Spec. Rep. 91-4, 33 p.

Collard, S.B. and C.N. D'Asaro. 1973. Benthic invertebrates of the eastern Gulf of Mexico. Sect. 3G. In: *A summary of knowledge of the eastern Gulf of Mexico* (J.I. Jones et al., eds.). State Univ. Sys. Florida, Inst. Oceanogr.

Day, J.W. et al. 1989. *Estuarine ecology*. John Wiley & Sons, New York, 558 p.

Durako, M.J., R.C. Phillips and R.R. Lewis III (eds.). 1987. *Proceedings of the symposium on the subtropical-tropical seagrasses of the southeastern United States*, 12 August 1985. Florida Mar. Res. Publ., Florida Dept. Nat. Resour. 42, 209 p.

Galtsoff, P.S. (ed.). 1954. *Gulf of Mexico, its origin, waters and marine life*. U.S. Fish Wildl. Ser., Fish. Bull. 64.

Godfrey, R.K. and J.W. Wooten. 1979. *Aquatic and wetland plants of the southeastern United States: Monocotyledons*. Univ. Georgia Press, Athens, 712 p.

Greening, H.S. and R.J. Livingston. 1982. Diel variation in the structure of seagrass associated epibenthic macroinvertebrate communities. *Mar. Ecol. Prog. Ser.* 7:147-156.

Hackney, C.T., S.M. Adams and W.H. Martin (eds.). 1992. *Biodiversity of the southeastern United States: Aquatic communities*. John Wiley & Sons, New York, 779 p.

Haddad, K.D. and B.A. Harris. 1986. A Florida GIS for estuarine management. *Am Soc. Photogramm. Remote Sensing and Am Cong. Survey Mapping. Tech. Pap., Geog. Info. Syst.* 3:2-11.

Hurlburt, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54(2):187-211.

- Iverson, R.L. and H.F. Bittaker. 1986. Seagrass distribution and abundance in eastern Gulf of Mexico waters. *Estuar. Coast. Shelf Sci.* 22:577-602.
- Kennedy, V.S. 1982. *Estuarine comparisons*. Academic Press, New York, 709 p.
- Kennish, M.J. 1990. *Ecology of estuaries*. Vol. 2. Biological aspects. CRC Press, Boca Raton, Florida, 391 p.
- Kusler, J. and M.E. Kentula (eds.). 1989. *Wetland creation and restoration: the status of the science*. Vol. I: Regional reviews; Vol. II: Perspectives. EPA/600/3-89/038a,b, Environ. Res. Lab., Corvallis, Oregon.
- Larkum, A.W.D. et al. (eds.). 1989. *Biology of seagrasses*. Aquatic Plant Studies 2. Elsevier Press, New York, 841 p.
- Leber, K.M. and H.S. Greening. 1986. Community studies in seagrass meadows: a comparison of two methods for sampling macroinvertebrates and fishes. *Fish. Bull.* 84(2):443-450.
- Lewis, F.G. and A.W. Stoner. 1981. An examination of methods for sampling macrobenthos in seagrass meadows. *Bull. Mar. Sci.* 3(1):116-124.
- Livingston, R.J. 1977. Time as a factor in biomonitoring estuarine systems with reference to benthic macrophytes and epibenthic fishes and invertebrates. Pp. 212-234. In: *Biological Monitoring of Water and Effluent Quality* (J. Cairns et al., eds.). Am. Soc. Test. Materials, Spec. Tech. Publ. 607.
- Livingston, R.J. 1984. The relationship of physical factors and biological response in coastal seagrass meadows. *Estuaries* 7(4A):377-390.
- Magoon, O.T. et al. (eds.). 1985 (et seq.). *Coastal Zone '85 (et seq.)*. Proc. symp. coast. ocean manage., Am. Soc. Civil Eng., New York.
- Mann, K.H. and J.R.N. Lazier. 1991. *Dynamics of marine ecosystems*. Blackwell Scientific Publ., Boston, 466 p.
- Myers, R.L. and J.J. Ewel (eds.). 1990. *Ecosystems of Florida*. Univ. Central Florida Press, Orlando, 765 p.
- National Research Council. 1990. *Managing troubled waters: the role of environmental monitoring*. Mar. Bd., Comm. Eng. Tech. Syst. Natl. Acad. Press, Washington, D.C., 136 p.
- Ogren, L.H. and H.A. Brusher. 1977. The distribution and abundance of fishes caught with a trawl in the St. Andrew Bay system, Florida. *Northeast Gulf Sci.* 1(2):83-105.
- Phillips, R.C. 1960. Observations on the ecology and distribution of the Florida seagrasses. *Florida St. Bd. Conserv. Prof. Pap. Ser. 2*, 72 p.
- Phillips, R.C. and C.P. McRoy (eds.). 1980. *Handbook of seagrass biology: An ecosystem perspective*. Garland STPM Press, New York, 345 p.
- Pruitt, B.A. 1988. St. Andrews Bay seagrass study, Panama City, Florida, June 16-19, 1988. Unpubl. Interim Rep. Mar. Wetlands Unit, USEPA.
- Robblee, M.B. and J.C. Zieman. 1984. Diel variation in the fish fauna of a tropical seagrass feeding ground. *Bull. Mar. Sci.* 34(3):335-345.
- Routledge, R.D. 1980. Bias in estimating the diversity of large, uncensused communities. *Ecol.* 6(2):276-281.

- Seaman, W., Jr. (ed.). 1985. Florida aquatic habitat and fishery resources. Florida Chapter, Am. Fish. Soc., 543 p.
- Shaffer, R.N. 1992. A bibliography of research on St. Andrew Bay and the nearby coastal waters -- preliminary draft. Unpubl., 34 p.
- Soule, D.F. and G.S. Kleppel (eds.). 1988. Marine organisms as indicators. Springer-Verlag, New York, 327 p.
- Sullivan, M.J. and C.A. Moncreiff. 1990. Edaphic algae are an important component of salt marsh food-webs: evidence from multiple stable isotope analyses. Mar. Ecol. Prog. Ser. 62:149-159.
- Thayer, G.W., D.A. Wolff, and R.B. Williams. 1975. The impact of man on seagrass ecosystems. Am. Sci. 63:288-296.
- Turner, R.E. 1990. Landscape development and coastal wetland losses in the northern Gulf of Mexico. Am. Zool. 30:89-105.
- U.S. General Accounting Office. 1988. Water pollution: Stronger enforcement needed to improve compliance at federal facilities. GAO/RCED-89-13, 78 p.
- Ward, D.B. and C.J. Chapman. 1973. Military herbicide driftage and Florida vegetation. Florida Sci. 36(2-4):110-122.
- Wolff, D.A., J.A. Reidenauer and D.B. Means. 1988. An ecological characterization of the Florida Panhandle. U.S. Fish Wildl. Serv., Miner. Manage. Serv. FWS Biol. Rep. 88(12); MMS OCS Study 88-0063, 276 p.
- Young, D.K. and M.W. Young. 1978. Regulation of species densities of seagrass-associated macrobenthos: Evidence from field experiments in the Indian River estuary, Florida. J. Mar. Res. 36(4):569-593.
- Zieman, J.C. 1987. A review of certain aspects of the life, death, and distribution of the seagrasses of the southeastern United States 1960-1985. Pp. 53-88. In: Proceedings of the symposium on tropical-subtropical seagrasses of the southern United States (M.J. Durako, R.C. Phillips, and R.R. Lewis III, eds.). Florida Dep. Nat. Resour. Mar. Res. Publ. 42.
- Zieman, J.C. and R.T. Zieman. 1989. The ecology of the seagrass meadows of the west coast of Florida: A community profile. U.S. Fish Wildl. Serv. Biol. Rep. 85(7.25):1-155.
- Zimmerman, M.S. and R.J. Livingston. 1976. Seasonality and physicochemical ranges of benthic macrophytes from a north Florida estuary (Apalachee Bay). Cont. Mar. Sci. 20:33-45.

**A PRELIMINARY STUDY OF THE
WEATHERING OF JET FUELS IN SOIL
MONITORED BY SFE WITH GC ANALYSIS**

**Larry E. Gerdorn, PhD
Associate Professor
Division of Natural Science**

**Mobile College
P.O. Box 13220
Mobile, Alabama 36663-0220**

**Final Report for:
Summer Research Program
Air Force Civil Engineering Laboratory
Tyndall Air Force Base, Florida**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

September 1992

**A PRELIMINARY STUDY OF THE
WEATHERING OF JET FUELS IN SOIL
MONITORED BY SFE WITH GC ANALYSIS**

**Larry E. Gerdorn, PhD
Associate Professor
Division of Natural Science
Mobile College**

Abstract

An offline supercritical-fluid extraction method was further developed for soil samples containing jet fuels. Supercritical-fluid carbon dioxide was used for extraction at 80 °C and 300 atm. This method is rapid with an extraction time of 20 minutes or less and it appears to be nearly quantitative for JP-4. Laboratory extracts were analyzed by GC with an FID detector. Modeling studies designed to determine how jet fuel components weather at a spill site are incomplete at this time.

**A PRELIMINARY STUDY OF THE
WEATHERING OF JET FUELS IN SOIL
MONITORED BY SFE WITH GC ANALYSIS**

Larry E. Gerdorn, PhD

INTRODUCTION

The isolation of organic compounds from complex matrices is a limiting step in their analysis. A new approach to this problem is provided by analytically scaled supercritical-fluid extraction or SFE. A supercritical-fluid is a gas that has been heated and compressed beyond its critical point. The supercritical state has properties intermittent between those of a liquid or a gas. Thus SFE is an isolation method for organic compounds that employs the special properties of supercritical-fluids. Supercritical carbon dioxide has been found to be an excellent solvent for many organic pollutants including polychlorinated biphenyls(PCB's), polyaromatic hydrocarbons(PAH's), and various herbicides.¹

SFE has several advantages over the traditional methods of isolation such as solvent extraction or thermal analysis. Solvent extraction is generally time consuming and usually generates large amounts of waste solvents that must be safely disposed. In addition, concentration and purification of extracts is necessary before subsequent analysis. Thermal methods are limited by the stability of the matrix and the analyte in question. SFE has none of the above limitations.²

DISCUSSION OF PROBLEM

The Air Force must store large quantities of jet fuel for use by its aircraft. Accidental discharge of fuels into the environment is a normal part of storage and transport. When these spills occur it is important to know the identity of the jet fuel, the level of contamination, and if possible the approximate time when the discharge occurred from a storage tank. It has already been established that each Jet Fuel has its own characteristic GC chromatogram that allows for a finger print type of identification of any given fuel(Figure 1).³ In cases where Air Force fuels are stored on commercial air bases the ability to identify fuels has been helpful in determining liability. Furthermore, if the weathering of fuels in the environment could be modeled it might be possible to establish when a spill occurred and thus help in estimating the extent of contamination.

An SFE method would be helpful in the analysis of jet fuels in soil because of the above mentioned advantages of this method. SFE had already been used to develop methods for the analysis of coal tar contaminated soils.^{4,5} It therefore seemed reasonable that an SFE method could be

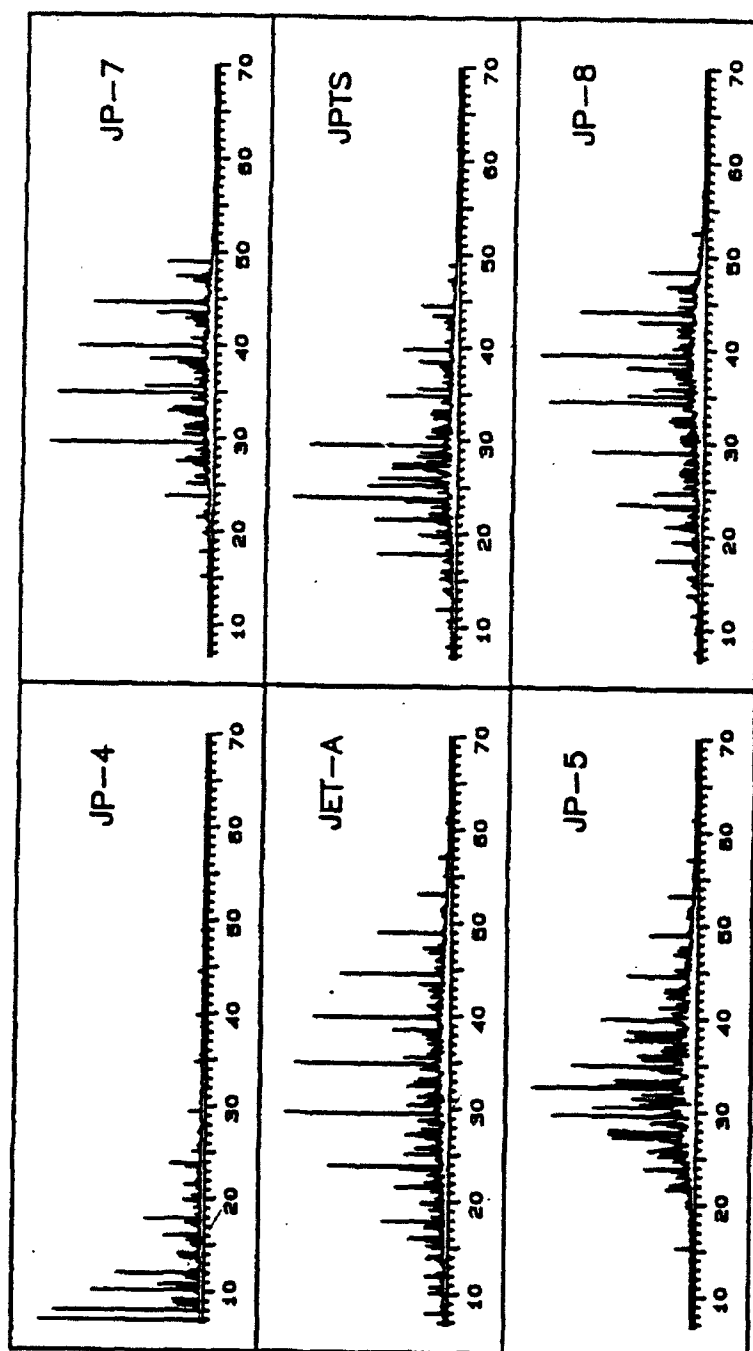


FIGURE 1. Chromatograms of Typical Jet Fuels.

developed to qualitatively and quantitatively analyze Jet Fuels from soil samples.

This project involved the development of an SFE extraction procedure for jet fuel in soils, the installation of a remote sensor weather station to gather information on weather conditions, and preliminary data on the changes with time of JP-4 in a soil matrix.

METHODOLOGY

SFE has several parameters that can be used to enhance recovery of samples including: the choice of gas to be used as the supercritical fluid, the pressure for extraction, the temperature for extraction, and the use of solvent modifiers in the extraction fluid. Since supercritical carbon dioxide has been shown to be a good solvent for most PAH's it was chosen as the extraction fluid.²

A Suprex SFE/50 extractor was used in this study. The temperature for extraction was set at 80 °C to volatilize any fuel components in the sample and the pressure for extraction was set at 300 atm. The technical support personnel at Suprex Corporation had recommended 400 atm as the extraction pressure but the SFE/50 is designed for a maximum pressure of 360 atm. The lower pressure did not prove to be a problem in these studies. The extraction procedure began with a 10 minute static extraction followed by a 30 minute dynamic extraction into a suitable solvent. Before extraction, 100 µL of the trapping solvent which contained d-10 ethylbenzene(2.50 ppm v/v) was added to the soil sample as a solvent modifier and for use as an internal standard for quantitative analysis. Initially extracts were collected in acetonitrile but later this solvent was changed to hexane which proved to be a better trap for the more volatile fuel components. For work involving standard solutions a 2 mL extraction vessel was used.

Standard solutions were prepared using benzene, toluene, isopropylbenzene, undecane, tridecane, and pentadecane dissolved in 10.00 mL of trapping solvent. These standard solutions contained from 1 to 25 µL of each of these standard compounds along with 25 µL of d-10 ethylbenzene added as an internal standard. These compounds were chosen because they are a good cross section of the components normally found in jet fuels in terms of boiling points. A 100 µL spike of each standard solution was added to a previously extracted sample of beach sand. A vial containing 1 mL of trapping solvent which was placed into a cooled aluminum block was used for sample collection.

The sample was extracted and the extract solution was analyzed by gas chromatography(GC) with a flame ionization detector(FID). The HP 5890 GC was set for splitless injection. The injection port temperature was 250 °C and the column temperature was initially set at 40 °C for 4 minutes. The temperature was then ramped to 350 °C at 10 degrees /minute. The temperature was then held constant for an additional 10 minutes.

A sample containing 100 µL of standard solution in 1 mL of hexane was analyzed by GC/MS

to identify the retention times for each component. Standard samples for each concentration were similarly prepared and analyzed by the GC method. The ratio of peak area for d-10 ethylbenzene to each component was used to quantify results.

In conjunction with the development of this method outdoor bins containing beach sand were prepared for the weathering study. Each bin contained 78.0 Kg(172 pounds) of dried and sifted beach sand. The sand used had been dried for 24 hours at 60 °C then sifted through a series of screens(4.75 mm, 2.00 mm, 600 µm, 600 µm, and 4.25 µm) to remove the bulk organic material. Sand samples were stored in closed 30 gallon plastic trash cans until needed. Jet fuel was spiked onto the sand and sand samples were taken at regular time intervals using a sampling device built in the machine shop by Mr. Bob Dunn. The samples were stored in capped glass bottles and analyzed as soon as possible by SFE and GC.

A weather station was setup that monitored air temperature, soil temperature in each test bin, barometric pressure, and relative humidity. This data was taken continuously every 20 minutes during experiments and could be down loaded to a computer inside the research lab by use of a connecting cable that was installed.

RESULTS

Figure 2 shows the result of extractions using standard solutions with acetonitrile as the trapping solvent. The gas chromatograms for these compounds showed that benzene and toluene were not being trapped as efficiently as the other components in the test solution. It was decided to use hexane as the trapping solvent.

Figure 3 shows the results of the study using hexane as the trapping solvent. The more volatile components of the test samples still did not seem to be trapped very efficiently. It was decided to monitor the amount of CO₂ gas that was passing through the extraction cartridge by using the density of the supercritical-fluid and the flow rate. The extraction at this point was not time dependent but rather volume dependent and extraction was halted after three cell volumes of fluid had passed through the extraction cell. Figure 4 shows the gas chromatograms of an extracted sample of JP-4 fuel and a pure sample of JP-4 fuel. It is clear that there is very little loss of the volatile components using this method. Time did not permit the resampling of all the standard solutions to verify this new method. However, the similarity in peak heights for these two chromatograms suggests that the extraction efficiency is nearly 100% using this method.

Two experiments using the sample bins were performed. In each case 80.00 mL of fresh JP-4 was added to the bin and stirred into the sand. A sample was immediately taken from the bin. Subsequent samples were taken every hour for at least the first nine hours. To increase the detection

ACETONITRILE EXTRACTION CURVES FOR SIX COMPOUNDS

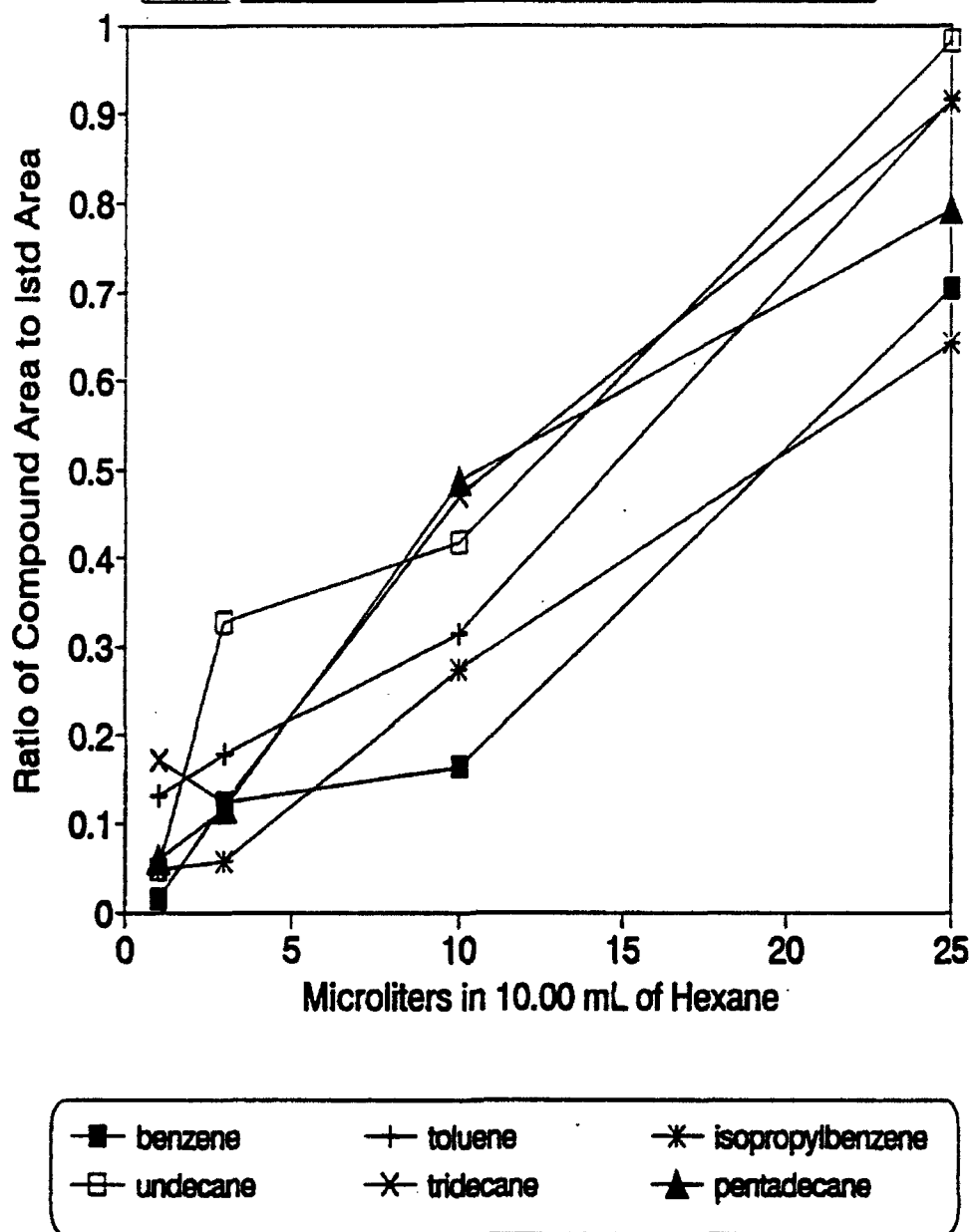


FIGURE 2.

HEXANE EXTRACTION CURVES FOR FIVE COMPOUNDS

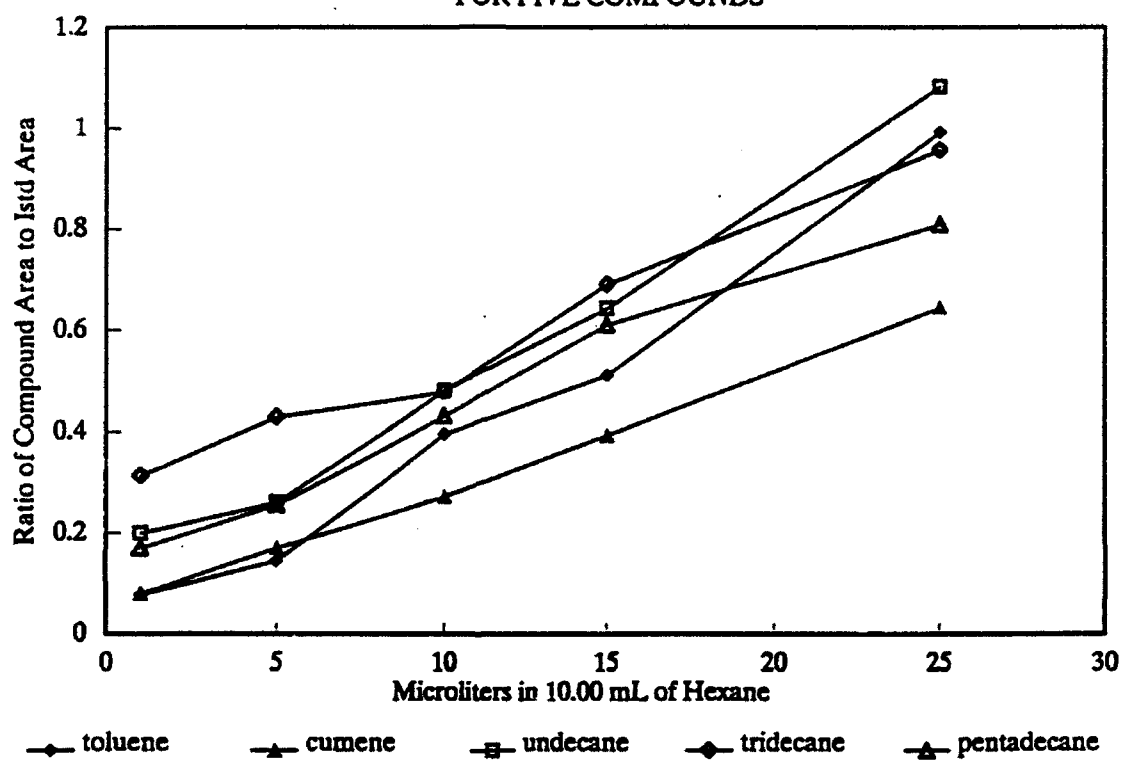
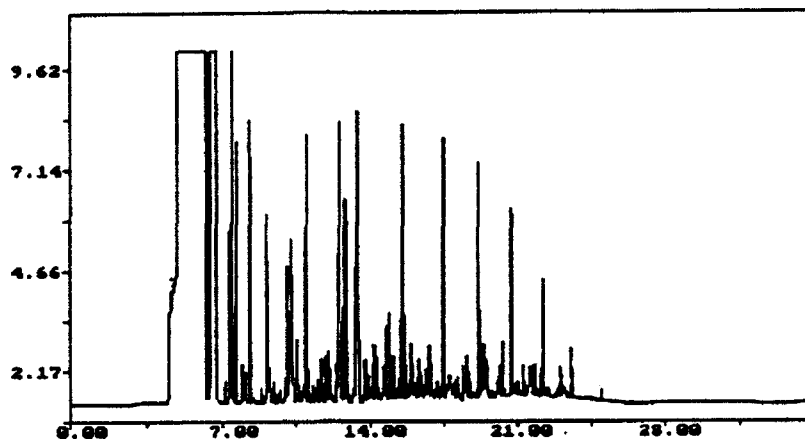
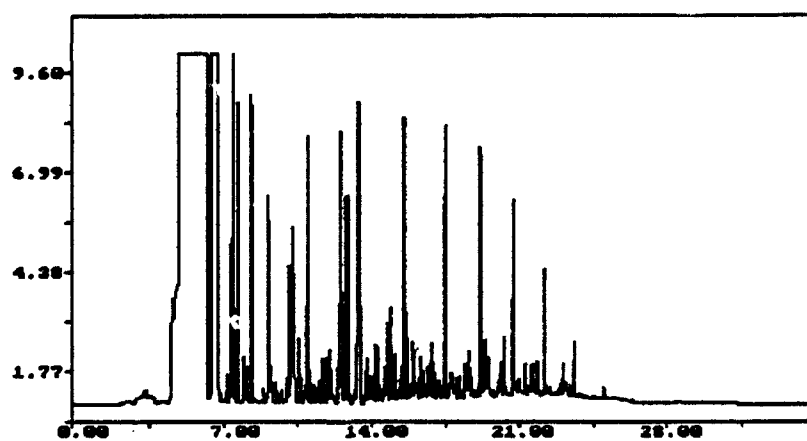


FIGURE 3.



Extracted Sample



Standard Jet Fuel Sample

FIGURE 4.

limits of the method extraction vessels of 5 mL and 12 mL were tested. This allowed larger samples of sand to be sampled while the amount of trapping solvent remained the same. It was found that within 36 hours very little of the JP-4 remained based on the GC chromatograms of extracted samples. This was very disappointing since the idea was to mimic accidental spills in the environment from underground storage tanks. Figure 5 shows the concentration changes with time for four different components of the fuel. The spikes in the concentration curve are not completely explainable at this point but probably relate to the inability to thoroughly mix the jet fuel with the sand. Further studies of this type might use a cement mixer or similar device to mix the jet fuel into the sand.

CONCLUSIONS

A quantitative and qualitative method for detecting jet fuels of interest to the Air Force was developed using SFE offline with analysis by a GC with an FID detector. This method is quick with extraction times of about 20 to 25 minutes. The GC analysis requires about an hour for each sample.

The effect weather plays on fuel samples is yet to be determined. It is assumed that temperature, pressure, and relative humidity will be important considerations. Further studies may give attention to the effect the water table plays in fuel spills. It is known that in cases where fuel has accidentally been released into the environment that the fuel percolates down through the soil until it reaches the water table. It then tends to spread out over the surface of the water since jet fuel is less dense than water and does not tend to mix with it. Such studies should be helpful in the cleanup of environmentally hazardous sites.

I wish to thank the Air Force Office of Scientific Research for supporting this study, the officers and staff of Tyndall Air Force Base for their assistance and advice in completing this work, and Mr. Robert Kuhn for helping in the analysis of samples.

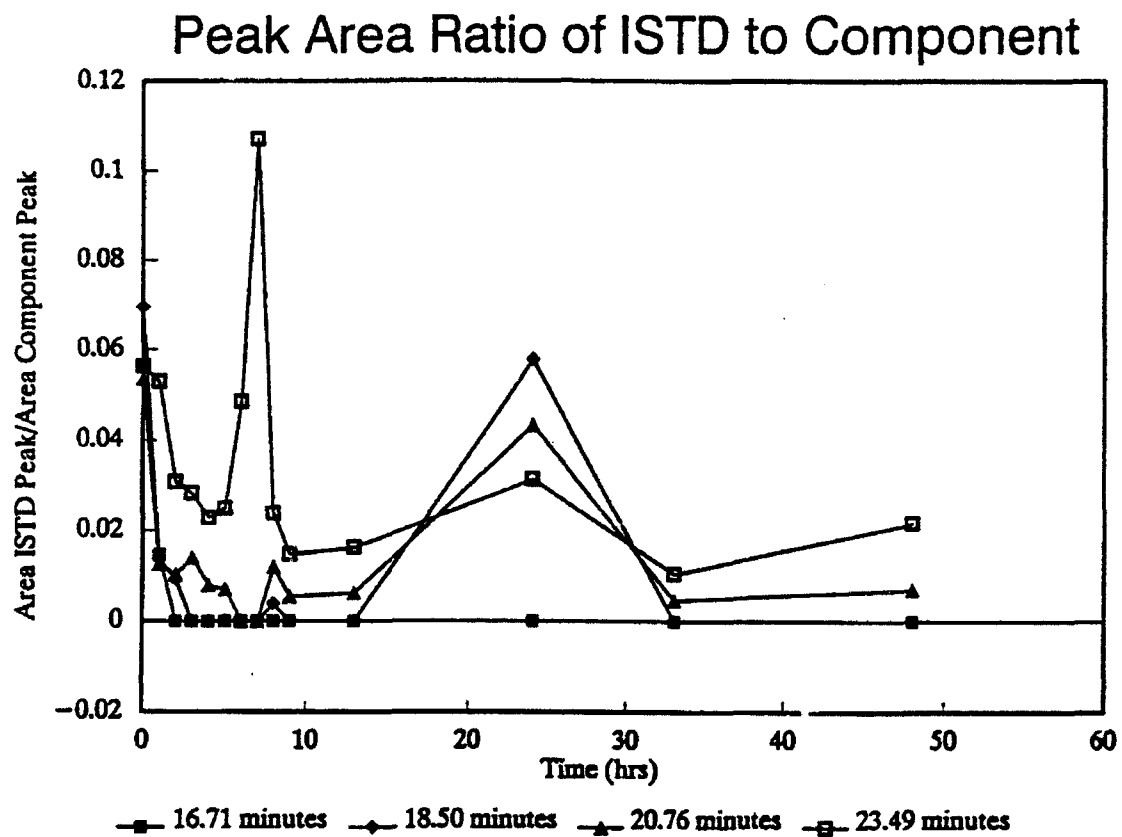


FIGURE 5.

REFERENCES

1. V. Lopez-Avila, N. S. Dodhiwala, and W. F. Beckert. "Supercritical-Fluid Extraction and Its Application to Environmental Analysis." J. Chrom. Sci., 28, 468-476(1990).
2. M. Lohleit, R. Hillmann, and K. Bächmann. "The Use of Supercritical-Fluid Extraction in Environmental Analysis." Fresenius J. Anal. Chem., 339, 470-474(1991).
3. H. T. Mayfield. "Air Force Fuel Identification Studies."
4. B. W. Wright, C. W. Wright, and J. S. Fruchter. "Supercritical-Fluid Extraction of Coal Tar Contaminated Soil Samples." Energy & Fuels, 3(4), 474-480(1989).
5. X. Yu, X. Wang, R. Bartha, and J. D. Rosen. "Supercritical-Fluid Extraction of Coal Tar Contaminated Soil." Environ. Sci. Technol., 24, 1732-1738(1990).

PRELIMINARY NUMERICAL MODEL
OF GROUNDWATER FLOW AT THE MADE2 SITE

Donald D. Gray
Associate Professor
Department of Civil Engineering

West Virginia University
Morgantown, WV 26506-6101

Final report for:
Summer Research Program
Civil Engineering Support Agency

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

September 1992

PRELIMINARY NUMERICAL MODEL
OF GROUNDWATER FLOW AT THE MADE2 SITE

Donald D. Gray
Associate Professor
Department of Civil Engineering
West Virginia University

Abstract

The Air Force has embarked on a program of environmental restoration which includes many sites where the groundwater has been contaminated by jet fuel or solvents. In order to design effective groundwater remediation plans, it is necessary to be able to predict how dissolved and immiscible pollutants will be transported by advection and dispersion. This capability is being developed by the Civil Engineering Support Agency. As part of this effort, the public domain groundwater flow program MODFLOW and its associated programs were applied to predict the groundwater flow at the site of the Macrodispersion Experiment 2 on Columbus Air Force Base in Mississippi. The program was used in the steady state mode with hydrologic conditions as they existed in October, 1990. Field data were used to estimate the boundary conditions and the hydraulic conductivity distribution. The net recharge to the aquifer was estimated from meteorological data. The importance of accurately specifying the recharge was illustrated by performing simulations with and without recharge. In order to interpret the head distributions predicted by MODFLOW, particle trajectories were calculated by the program MODPATH and plotted using MODPATH-PLOT. Among the conclusions of this study are that realistic simulations must be three dimensional and transient, and that our ability to visualize the results lags our ability to perform the simulations.

PRELIMINARY NUMERICAL MODEL
OF GROUNDWATER FLOW AT THE MADE2 SITE

Donald D. Gray

INTRODUCTION

In recent years the Air Force has devoted increased attention to the problems of environmental pollution which have been created by its activities over the last several decades. Among the most important and persistent of these problems are those associated with groundwater pollution. The Air Force has determined that there are a large number of individual locations where jet fuel or solvents have been spilled onto the ground, either by accident or deliberately. What was perhaps not realized then, but is now all too apparent, is that these substances often reach the water table and foul the groundwater.

As the Air Force prepares to deal with this groundwater pollution, it has become evident that only when one knows where the pollutants are and what their concentration is can a rational plan of remediation be assigned. An efficient remediation strategy requires a capability to predict where and when dissolved or immiscible pollutants will be transported by the groundwater flow. There are numerous computer programs which have been written to provide this capability. My assignment as an AFOSR Summer Faculty Fellow was to implement and evaluate some of these programs on the Civil Engineering Support Agency's workstations. In particular, I was asked to attempt to predict the transport of pollutants which had been observed in the Macrodispersion Experiment 2 (MADE2) at Columbus Air Force Base, Mississippi, in 1990-91.

METHODOLOGY

My first step was to inventory the set of groundwater computer programs which had been assembled before my appointment began. Table 1 lists these programs and some of their important characteristics. These particular numerical models had been selected because they were public domain programs which were more or less

well documented and widely used.

Table 1. Inventory of groundwater models on hand, June 1992.

MODEL, VERSION	DATE	FLOW?	TRANS- PORT?	D I M .	UNSAT.?	PRE.?	POST.?	METHOD
MODFLOW, 4.2	11/91	yes	no *	3	no	yes	yes	FD
HST3D, 1.5	2/92	yes	yes	3	no	no	no	FD
SWICHA, 5.05	2/91	yes	yes	3	no	no	no	FE
SUTRA, 0690-2D	6/90	yes	yes	2	partial	yes	yes	FD/FE
MOC, 3.0	11/89	yes	yes	2	no	yes	no	FD/MOC
Random Walk	81	yes	yes	2	no	no	no	FD/RW
SAMFT1D, 1.0	9/90	yes	yes	1	yes	?	?	FD/FE

FD = finite difference, FE = finite element, MOC = method of characteristics,
RW = random walk

* Particle tracking code MODPATH is available.

As Table 1 shows, all of these programs have the ability to predict groundwater flow. Except for MODFLOW, they also have the capability to predict pollutant transport by solving the advection-diffusion equation. Although MODFLOW lacks this capability, there is a companion program (MODPATH) which predicts steady state particle trajectories, thus essentially providing an upper bound for pollutant transport. Only MODFLOW, HST3D, and SWICHA are able to solve for three dimensional flows. SAMFT1D, although limited to one dimensional problems, can predict the motion of up to three immiscible phases; the others are single phase models. SUTRA and SAMFT1D have some capability to predict flow in the

unsaturated zone, but the other models are valid only in the saturated zone. The PRE.? and POST.? columns refer to the existence of preprocessor and postprocessor programs to assist the user in preparing the input file and displaying the output file. Although such auxiliary programs are not essential, they are extremely useful in practice.

Although all of these programs are written in standard Fortran, the versions on hand had been prepared for either mainframe or personal computers. None had been ported to the recently established RAVC Research Modeling Computer Network. This network is based on two similar Sun Sparcserver2 workstations running Versions 2 and 3 of the Unix-based Open Windows operating system. Getting these programs to run on the workstations proved to be more time consuming than had been anticipated.

The Macrodispersion Experiment 2 provided an almost ideal test case for transport modeling because it was one of the most intensively monitored groundwater experiments ever conducted; and because the Air Force, as a participant in the experiment, had access to the data. On the other hand, the site was found to be extremely heterogeneous, and so would provide a stiff test for any numerical model.

The MADE2 test site was located on Columbus Air Force Base, Mississippi, and covered an area about 300 m x 200 m. The site was covered primarily by weeds and brush and had about 2 m of relief. There were no streams or ponds on the site. The MADE2 investigators found that the upper layer of soil was a shallow alluvial terrace containing a saturated zone about 11 m thick underlain by a clay aquitard (Boggs, Young, Benton, and Chung; 1990). The soil consisted of poorly sorted to well sorted sandy gravel and gravelly sand with minor amounts of silt and clay. The aquifer was found to contain irregular lenses and layers having typical horizontal dimensions on the order of 8 m and typical vertical dimensions on the order of 1 m. Figure 1 shows the locations of the wells at which

hydraulic conductivity was measured at 15 cm depth increments using the borehole flowmeter method. In all 2187 conductivity measurements were reported. Most of the other relevant aquifer properties were measured with similar comprehensiveness. Hydraulic conductivity variations of up to 4 orders of magnitude were found in individual wells.

With such extreme heterogeneity, it was obvious that only a three dimensional model could provide realistic predictions. Of the trio of three dimensional models listed in Table 1, HST3D seemed to be the most suitable, but I did not believe that I could adapt it to the Sun workstations in the available time. Based on my previous experience, I chose to devote most of my efforts to using MODFLOW.

MODFLOW (McDonald and Harbaugh, 1988) is a U. S. Geological Survey (USGS) program which solves a block centered finite difference approximation to the groundwater flow equation on a variable cell size, three dimensional rectangular grid. MODFLOW allows for anisotropy so long as the grid axes are aligned with the principal directions of hydraulic conductivity. The name MODFLOW refers to the modular structure of the program in which each optional feature is coded as a separate subroutine. Clearly understandable coding and outstanding documentation make MODFLOW the benchmark for evaluating other Fortran programs, regardless of field.

MODFLOW is very flexible. It can solve either steady or transient cases and provides options for recharge, wells, streams, and other hydrologic features. Both confined and unconfined aquifers can be modeled. The algorithm allows the dewatering of layers during periods of water table decline, but the version I used could not handle rewetting due to a rising water table (a valuable capability which has been added to later versions). MODFLOW offers three independent solvers and has earned a reputation for robustness. As a result of its many desirable features, MODFLOW is perhaps the most widely used groundwater

flow code in the U. S.

As the popularity of MODFLOW has increased, a number of auxiliary programs have been written to facilitate its use. PREMOD is a preprocessor which simplifies the task of preparing the complex and often extensive input files. The results of a MODFLOW execution are tables of heads and related variables written to several output files, some in ASCII format and some in binary format. POSTMOD is a postprocessor which converts the binary files into ASCII format so that they can be imported into spreadsheet or plotting programs for further analysis.

MODPATH (Pollock, 1989) is a USGS program which uses the heads calculated by MODFLOW to calculate seepage velocities. Based on these seepage velocities, MODPATH determines the trajectories of inert, nondispersive particles. The number of particles and their initial positions are arbitrary, but the flow regime must be in steady state. These trajectories provide information on where pollutants move ignoring dispersion. MODPATH-PLOT is a companion program which can create several types of graphical displays of the MODPATH output files.

In order to run MODFLOW, the user must specify certain data. These include the grid geometry, the boundary and initial conditions, values related to the principal hydraulic conductivities for each cell, and the sources of water if any. For unsteady simulations, the storage coefficients must be specified for each cell. To use MODPLOT, the user must also give the porosity for each cell.

Based on the MADE2 results, I chose to model a horizontal area of 340 m x 165 m using 5 m x 5 m cells. Thus the grid had 68 rows and 33 columns. Nine layers were used to discretize the vertical thickness of the aquifer. The bottom boundary of the grid was idealized as an impermeable plane at an elevation of 51 m, and the lower eight layers were each 1 m thick. The upper layer was an unconfined layer whose bottom was at 59 m and whose upper boundary was the

computed water table elevation. Thus the thickness of the saturated zone in the upper layer was variable. Because the upper layer could not be allowed to dewater, its bottom had to be set lower than would otherwise be desirable. The total number of cells in the entire grid was 20,196.

Piezometric heads had been measured continuously in 15 monitoring wells scattered over the site. In addition, manual measurements were made at about 48 wells approximately once a month during MADE2 in order to obtain detailed "snapshots" of the flow. Figure 2 shows the hydrograph measured at one of the continuously monitored wells. The squares indicating the "snapshot" data are in good agreement with the continuous record. From day 0 (June 26, 1990) until about day 180, the head declined less than 1 m in a smooth fashion. After that date the variation was much larger and more erratic. These results, which were typical of the continuously monitored wells, showed that a fully transient simulation would be essential for the later portion of the test. As a first step, however, it appeared that a steady state simulation would be adequate to predict the regime of October 15, 1990 (day 83).

The October 15, 1990, "snapshot" provided piezometric heads at 45 points scattered around the site. The kriging option in the commercial program SURFER was used to interpolate head values at the midpoint of each cell of the computational grid. All of the heads were assumed to represent water table elevations, and head variations with depth were ignored. These krigged heads were input to MODFLOW. The heads in the cells around the boundary of the grid were specified to be boundary values and were held constant during the computations. The heads in the interior cells were merely used as initial guesses and were changed by the calculations. The Environmental Protection Agency (EPA) program GEOPACK (Yates and Yates, 1990) was less satisfactory than SURFER for this application.

Although borehole hydraulic conductivity tests were reported at 67 wells, there

was time to reduce data from only six wells. It is clear that by using all of this data, more accurate simulations can be performed. At each well, horizontal conductivities were reported for vertical layers with thicknesses of about 15 cm. These data were used to calculate effective horizontal conductivities for each layer of the grid. Values ranged from 10^{-7} to 1.2×10^{-3} m/s. Where data for the top or bottom layers were not reported, the value for the vertically adjacent cell was used. The values derived for each well were assigned to all the cells in the nearest rows of the grid. Thus the values from well K49 were used for rows 1 through 18, the values from well K40 were used for rows 19 through 35, and so on. The hydraulic conductivities were entered directly for the unconfined top layer, but had to be converted to transmissivities for the lower layers. The layers were assumed to be horizontally isotropic.

Equivalent vertical conductivities between vertically adjacent nodes were based on the equivalent horizontal conductivities of the cells weighted by the node to cell boundary distances. This procedure assumes that each soil layer is locally isotropic, and that all anisotropy is due to layering. The node to boundary distance for the unsaturated top cell was arbitrarily assumed to be 0.5 m. These conductivities were converted to leakances for input to MODFLOW. The values ranged from 10^{-7} to 7.8×10^{-4} s⁻¹.

Except for flow through the lateral boundaries (the lower boundary was assumed impermeable), the only physical process by which water could enter or leave the site was through recharge. Because no lysimeter data was available, recharge was estimated from meteorological data. Daily temperature and precipitation data were measured at the CAFB weather station, less the 2 km from the test site. Figure 3 shows the precipitation during the test period. The nearest daily evaporation pan data was from State University, about 35 km distant, and was supplied by State Climatologist Dr. C. L. Wax. Missing evaporation data were estimated from the daily maximum temperatures using the empirical equation of Pote and Wax (1986). Figure 4 shows that the estimates based on this equation

fit the trend of the actual measurements with tolerable accuracy. Based on discussions with Dr. Wax, a pan coefficient of 0.8 was used to estimate the evapotranspiration. The net recharge was the difference between the daily precipitation and the evapotranspiration. The cumulative recharge is shown in Figure 5. The net recharge was averaged from September 1 through October 31, 1990 (days 67 through 127), to obtain the value of -2.34×10^{-4} m/s which was input to MODFLOW as the best estimate of recharge.

The porosity of each cell must be specified in order to allow MODPATH to calculate seepage velocities based on the heads output by MODFLOW. In spite of the seemingly exhaustive study which had been done of the MADE2 site, porosities had been measured at only 4 core holes. The 84 samples had a mean porosity of 0.32 and this value was used for every cell in the grid. Without additional field measurements, it will be difficult to improve upon this obviously inadequate model of the porosity field.

RESULTS

Steady state MODFLOW simulations of the MADE2 site using the October 15, 1990, conditions described above were run using the Sun workstations. The results were sensitive to the convergence parameters which were specified. With the Preconditioned Conjugate Gradient Solver, the following parameters gave satisfactory results: relaxation factor = 1.0, head change criterion for closure = 0.0005, and residual change for closure = 0.0005. One case was run using the recharge of -2.34×10^{-4} m/s, and a second case was run with a recharge of 0. By comparing these cases, the importance of correctly estimating the recharge rate can be seen. In principle, such sensitivity studies should be conducted for each input parameter, but the finite time available for such studies tends to limit sensitivity analyses to at most a few key parameters.

Both cases had a run time of about 2 minutes and required 63 iterations to converge. MODFLOW computes a global water budget which provides one measure of

the consistency of the calculations. Table 2 compares the water budgets for these simulations. It is interesting to note that 63 % of the outflow occurs by evapotranspiration when recharge is simulated. If a transient simulation were performed, this would cause a decline in the water table.

Table 2. Water budgets for steady state MODFLOW simulations of the MADE2 site for October 15, 1990, conditions.

Case:	with recharge	no recharge
Recharge rate:	$-2.34 \times 10^{-4} \text{ m/s}$	0
Inflow through boundary:	0.0018782 m ³ /s	0.0011460 m ³ /s
Outflow through boundary:	0.00067908 m ³ /s	0.0011277 m ³ /s
Loss through water table:	0.0011969 m ³ /s	0
Discrepancy:	0.12 %	1.61 %

In order to comprehend the meaning of the thousands of numbers produced by MODFLOW, it is virtually essential to reduce them to graphical form. Figures 6 through 9 were produced using MODPATH and MODPATH-PLOT. These particle trajectory plots reveal the highly three dimensional nature of the simulated flow. Figure 6 shows the trajectories of a line of particles which start in row 48 in the top layer in the simulation with recharge.. The particles converge toward the center of the grid as they move toward the north due to inflow through the east and west boundaries. The termination of the trajectories in the interior of the grid occurs because the particles have reached the water table and have been lost through evapotranspiration. This can be seen more clearly in Figure 7 which shows the trajectories of a vertical line of particles which start in column 10, row 48. These trajectories have been projected onto a vertical plane. The particle which starts in the top layer is seen to leave through the upper boundary. The other trajectories terminate where the particles flow

through the western boundary. The strongly three dimensional nature of the flow is caused primarily because the particles tend to flow through regions of higher hydraulic conductivity.

Figure 8 shows the trajectories of particles which started in row 48 in the top layer in the no recharge simulation. Since there is no evapotranspiration, the trajectories all extend to the lateral boundaries. Compared to Figure 6, the degree of convergence is much greater. The bifurcation of the trajectories near the northwest corner of the grid is caused by a region of higher head on the boundary. The refraction of the trajectories is caused by abrupt changes in the hydraulic conductivity. Figure 9 shows the vertically projected trajectories of particles which started in a vertical line of cells at row 48, column 10. The three dimensional nature of the flow remains pronounced; but unlike Figure 7, all of the trajectories terminate through the western boundary.

These plots are more informative than reams of tables, but they are only a small fraction of the plots which could be made. It is obvious that one of the major bottlenecks in groundwater modeling is visualizing the results. The outstanding graphical capability of the Sun workstations should prove invaluable in this regard.

It is important to realize that these simulations are not good representations of the actual MADE2 flows due to the many assumptions which were introduced in order to obtain results in the allotted time. Perhaps the most restrictive assumption is that the calculations are for steady state. Although the conditions at the MADE2 site were fairly steady during the first 180 days of the experiment, this period was far too short to reach steady state. Even if the flow field was steady, it did not remain steady long enough for the particles to trace out their entire predicted trajectories. For believable results, transient simulations are a necessity. This can be accomplished, but the increase in effort required will be substantial. In addition, it must be remembered that

MODPATH can calculate only steady state trajectories. Other limitations on the present results include the coarse discretization of the hydraulic conductivity and the unrealistic specification of a uniform porosity.

CONCLUSIONS

1. Realistic simulations of groundwater flow and contaminant transport must be three dimensional and transient. Codes such as HST3D have these capabilities.
2. No matter how much field data is available, it is never as much as the modeler wants. Perhaps some form of tomography is the answer.
3. The ability to visualize and interpret the results lags far behind the ability to do calculations. The graphical capabilities of the Sun workstations must be fully exploited in the future.
4. Even programs which claim to be written in ANSI standard Fortran are not fully portable. There seem to be few groundwater programs which have been ported to the Sun workstations.
5. The kriging software used in this study was not entirely satisfactory.
6. A geographic information system (GIS) would be invaluable in integrating various classes of field data and simulation results.

ACKNOWLEDGEMENTS

I enthusiastically thank Dr. Tom Stauffer, Capt. Sean Childress, and Dr. Howard Mayfield of CESA for their assistance. I am grateful to Dr. C. L. Wax of the Mississippi State University Department of Geology for providing evaporation data and advice.

REFERENCES

J. M. Boggs, S. C. Young, D. J. Benton, and Y. C. Chung; 1990. Hydrogeological Characterization of the MADE Site, EPRI Topical Report EN-6915, Electric Power Research Institute, Palo Alto, California.

M. G. McDonald and A. W. Harbaugh, 1988. Techniques of Water-Resources Investigations of the United States Geological Survey, Chapter A1, A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, U. S. Government Printing Office, Washington, D. C.

D. W. Pollock, 1989. Documentation of Computer Programs to Compute and Display Pathlines from the U. S. Geological Survey Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, Open File Report 89-381, U. S. Geological Survey, Reston, Virginia.

J. W. Pote and C. L. Wax, 1986. Climatological Aspects of Irrigation Design Criteria in Mississippi, Technical Bulletin 138, Mississippi Agricultural and Forestry Experiment Station, Mississippi State University, Mississippi State, Mississippi.

S. R. Yates and M. V. Yates, 1990. Geostatistics for Waste Management: A User's Manual for the GEOPACK (Version 1.0) Geostatistical Software System, EPA/600/8-90/004, U. S. Environmental Protection Agency, Robert S. Kerr Environmental Research Laboratory, Ada , Oklahoma.

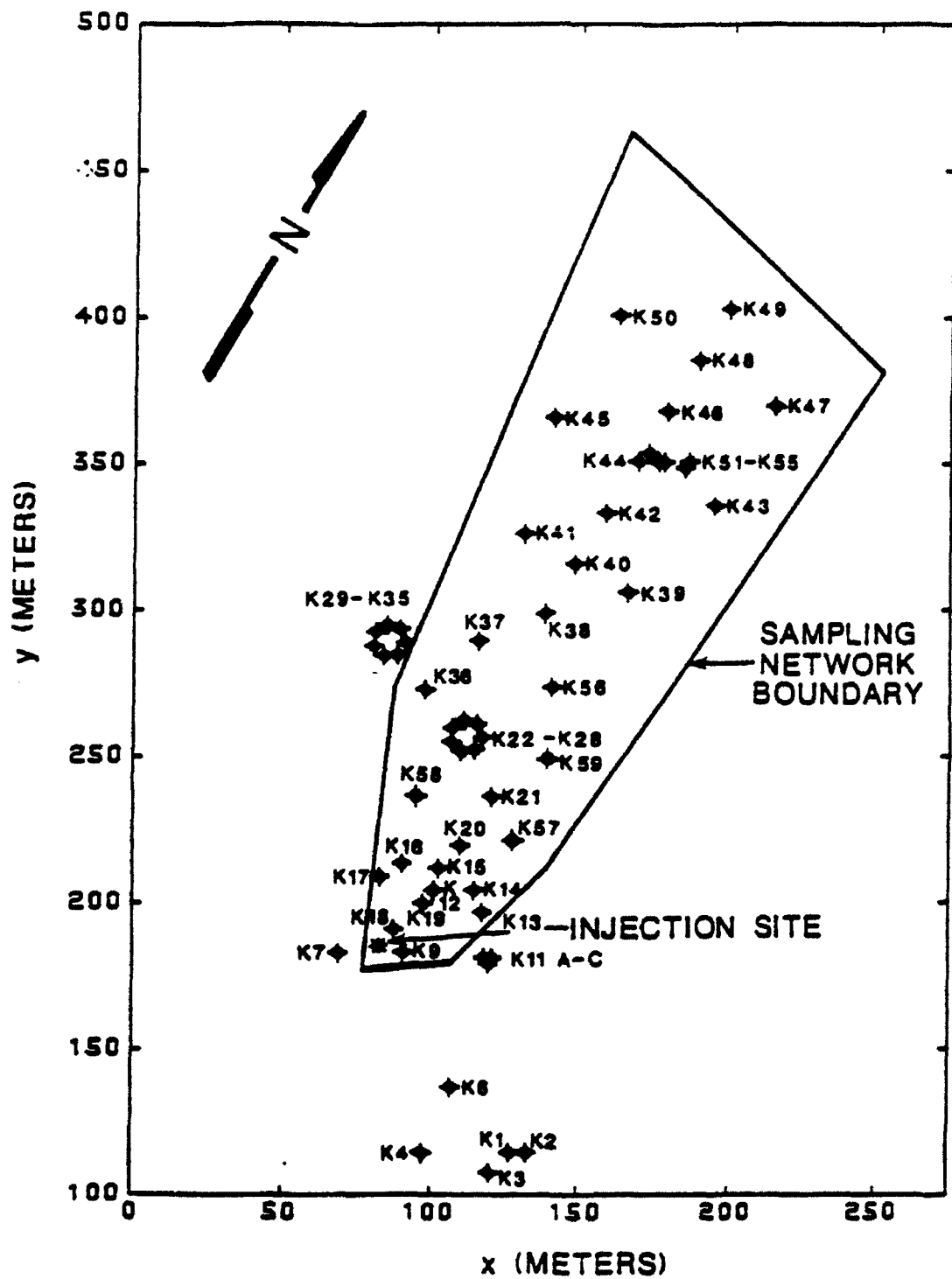


Figure 1. Borehole flowmeter test well locations at the MADE2 site.

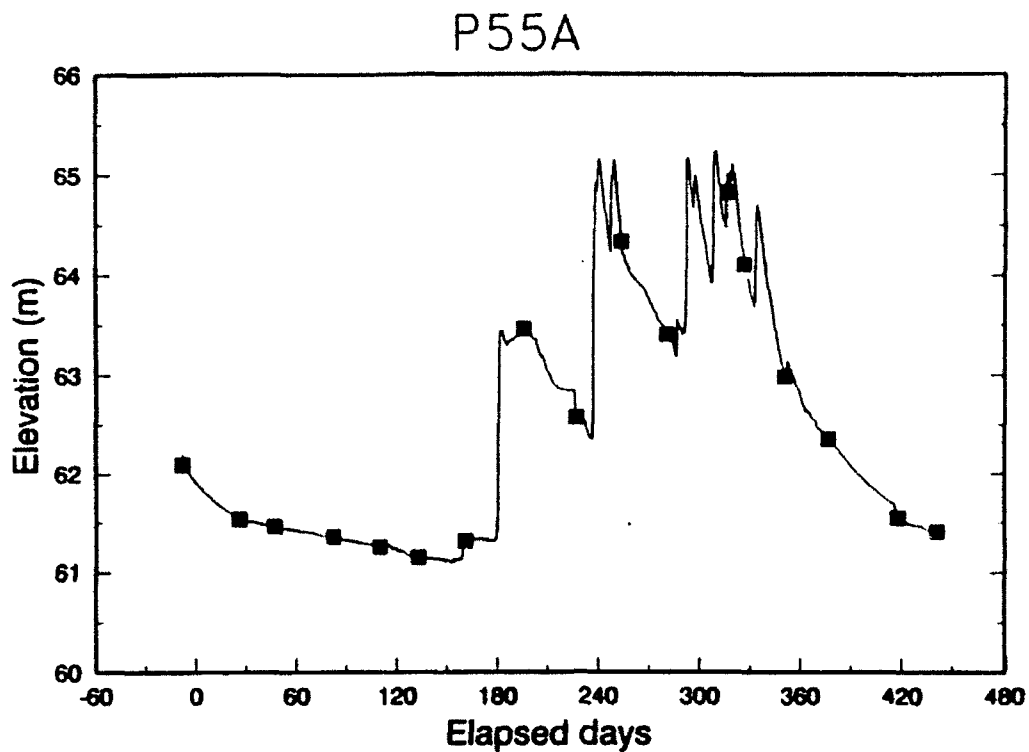


Figure 2. Hydrograph from typical monitoring well at the Made2 site. Squares represent manual "snapshot" measurements. Day 0 is June 26, 1990.

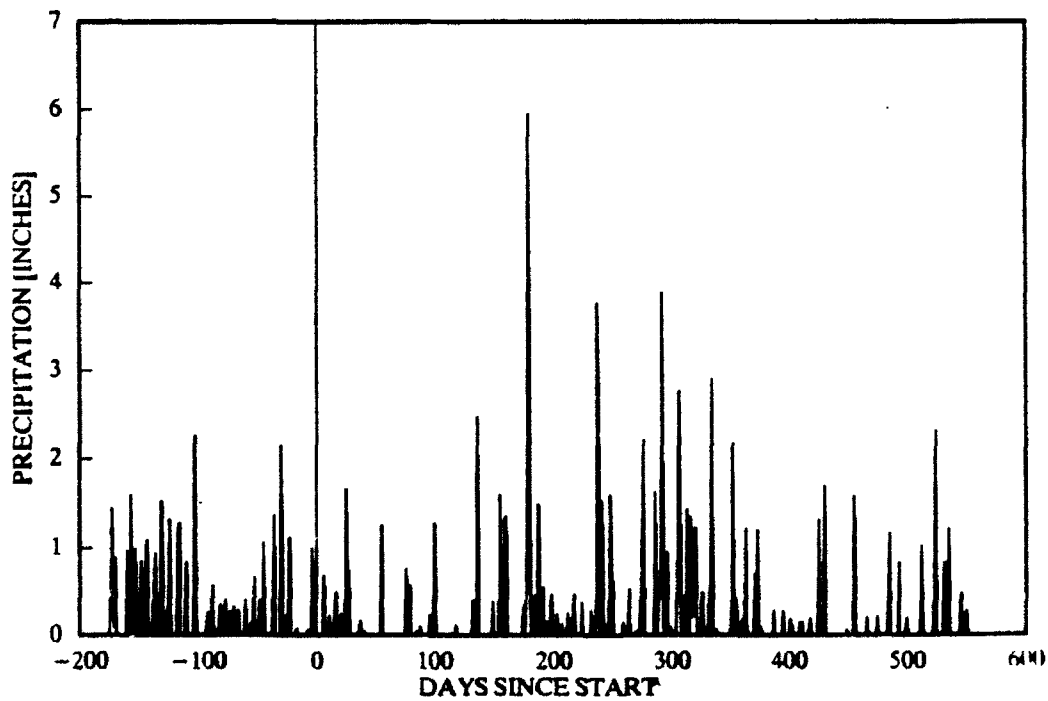


Figure 3. Daily precipitation at Columbus Air Force Base during MADE2.

ESTIMATED VS PAN EVAPORATION

MADE2 AT CAFB, MS, 1990-91

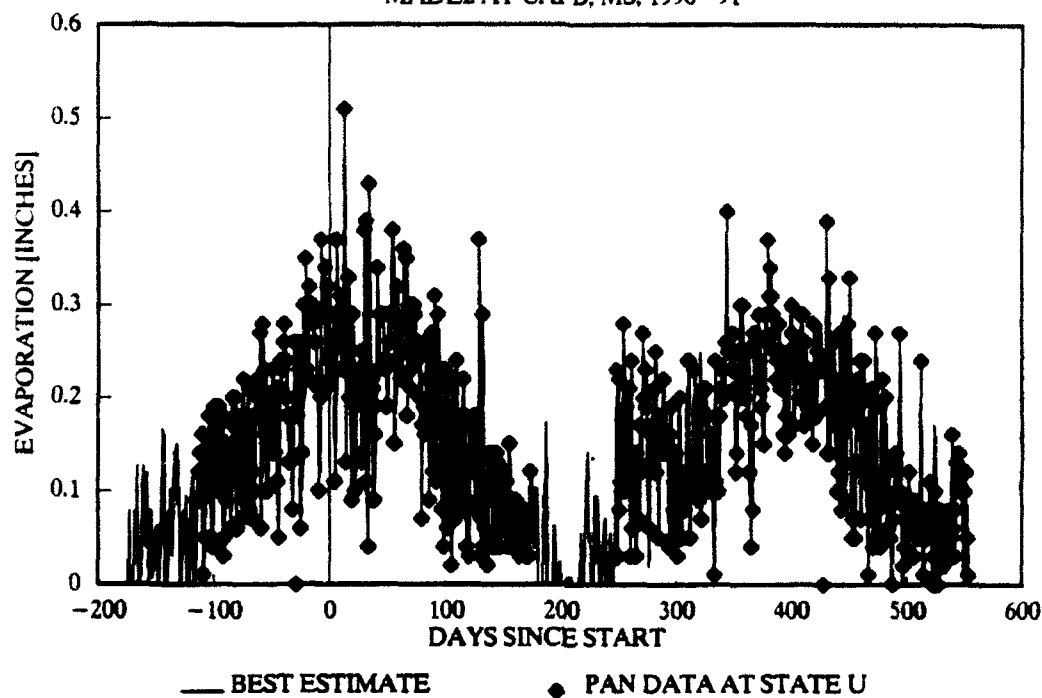


Figure 4. Daily pan evaporation during MADE2. Diamonds are actual measurements. Solid line uses Pote and Wax equation to complete the series.

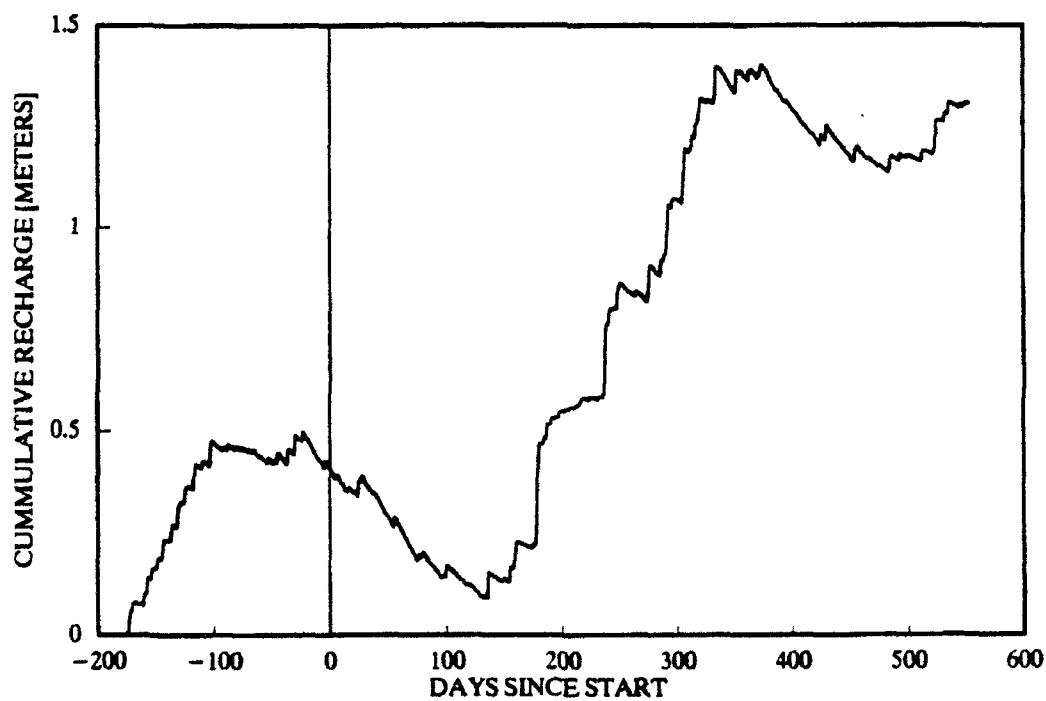
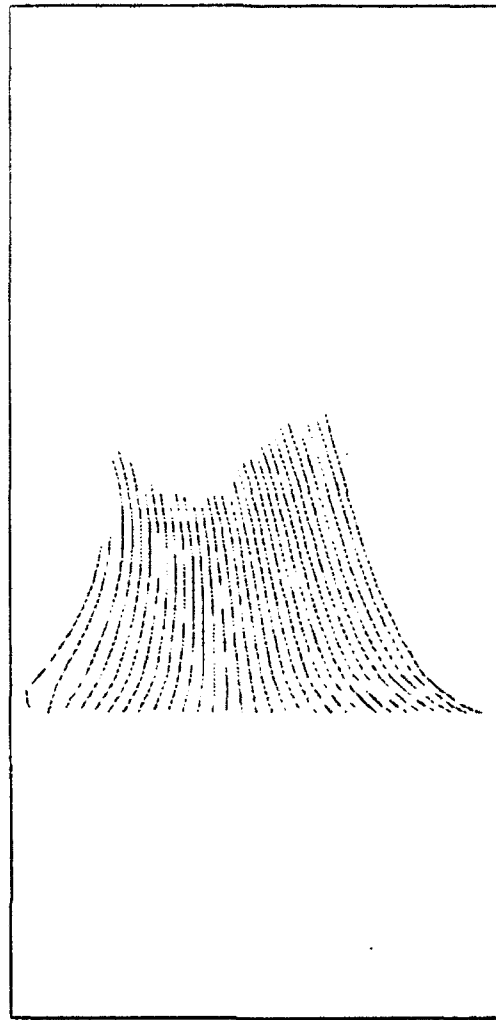


Figure 5. Cumulative recharge since January 1, 1990, at MADE2 site.



0 10 60 METERS

Figure 6. Steady state trajectories of particles starting in row 48, layer 1, with recharge.

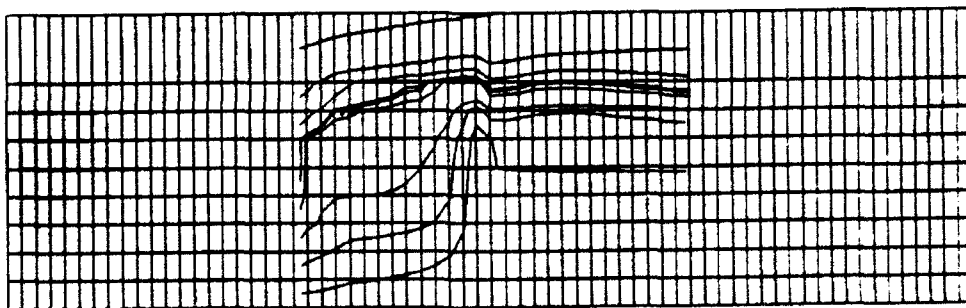
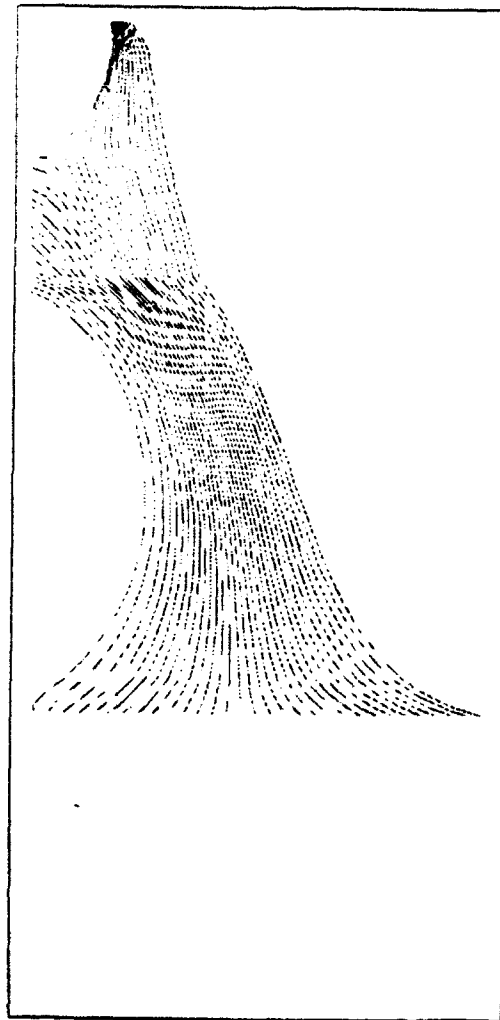


Figure 7. Steady state trajectories of particles starting in a vertical line at row 48, column 10, with recharge. The trajectories are projected onto a vertical plane. The vertical exaggeration is 10.



0 10 60 METERS

Figure 8. Steady state trajectories of particles starting in row 48, layer 1, no recharge.

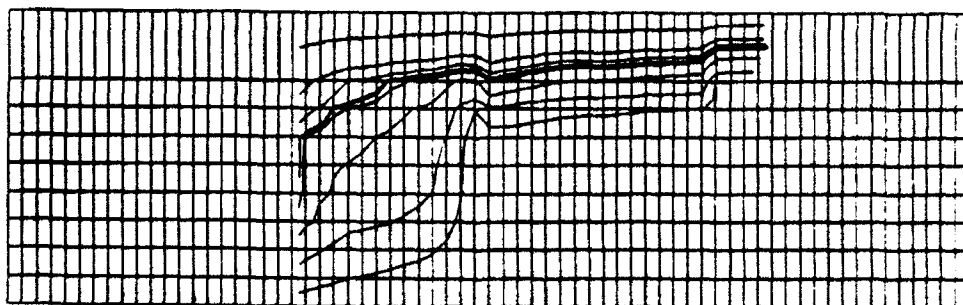


Figure 9. Steady state trajectories of particles starting in a vertical line at row 48, column 10, no recharge. The trajectories are projected onto a vertical plane. The vertical exaggeration is 10.

**DECISION ANALYSIS FOR SELECTION
OF HALON 1301 REPLACEMENT AGENTS**

**Charles J. Kibert
Associate Professor
Fire Research and Testing Center**

**University of Florida
FAC 101
Gainesville, Florida 32611-2032**

**Final Report for:
Summer Research Program
Air Force Civil Engineering Support Agency**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

September 1992

DECISION ANALYSIS FOR SELECTION
OF HALON 1301 REPLACEMENT AGENTS

Charles J. Kibert
Associate Professor
Fire Research and Testing Center
University of Florida

Abstract

The U.S. Air Force program to replace Halon 1301 in total flood applications with suitable compounds entails the requirement to set selection criteria that will provide technical and management decisionmakers with the basis for evaluating the alternatives. Classic decision matrices allow a single evaluator to both rate and provide weights for each criterion against the various alternatives. The methodology described in this paper allows multiple evaluators to rank the criteria in order to generate criteria weights. A symbolic scheme to state the relative importance of the criteria and a system for "collapsing" the rankings are described.

DECISION ANALYSIS FOR SELECTION OF HALON 1301 REPLACEMENT AGENTS

Charles J. Kibert

INTRODUCTION

The Montreal Protocol of 1987 and the Clean Air Act Amendments of 1990 have dictated that the class of chemicals known as chlorofluorocarbons (CFC's) be banned from production by January 1, 1995 and from use by January 1, 2000. Largely used as refrigerants, CFC's also play an important role as firefighting agents where clean fire suppression is important. In computer and communications facilities it is desirable to extinguish the various classes of fires that may occur with minimal disruption to ongoing operations and without adding to the damage caused by the fire suppression method. Use of water sprinklers, foams, or dry chemicals necessitate extensive and expensive cleanup operations after a fire. A subset of CFC's known as Halons has provided clean fire suppression capability for over two decades. Halon 1301 has been the clean agent most frequently employed to protect computer/communications facilities from fire damage and collateral damage that would be a function of the fire suppression method.

DISCUSSION

In another program to select a replacement agent for Halon 1211, used in flightline firefighting and in facility and aircraft portable firefighting units, a decision analysis matrix was utilized rather late in the program to provide a framework to justify the selection of the replacement agent (Figure 1.). Halon

1211 was compared to PFC-614 and HCFC-123 to determine which of the replacement agents should be selected for further testing. Criteria such as agent effectiveness, acute toxicity, system conversion costs, purchase cost, ozone depletion potential (ODP), and greenhouse warming potential (GWP) were used as the basis for agent evaluation. A few key features of this rating scheme can be note in Figure 1.

First, each criterion was treated identically in terms of weight. Fire suppression efficiency received the same weight or emphasis as toxicity and cost. For all practical purposes it could be said that no weighting scheme was utilized.

Second, the scoring of each alternative was accomplished by awarding three points for the alternatives with the best performance against each criterion and one point against each criterion showing the worst performance. The spread of points is not enough to significantly differentiate good performance from poor performance.

Third, several criteria are actually counted a number of times. The category entitled future regulatory phase-out since ODP, GWP, HCFC's, EPA SNAP approval are all connected with this topic. Consequently out of 16 criteria, four are counted against possible phase-out.

This example illustrates several of the pitfalls associated

with decision matrix methods. A set of clearly defined criteria are necessary in order to set up the decision matrix. Each criterion should appear once in the matrix and there should be no interdependence of criteria. This latter recommendation may not always be possible to implement. Nonetheless it should be utilized to the maximum extent possible. A weighting scheme should also be employed in order to indicate the relative priority of the criteria. The weights should be generated using input from several experts in the field in order to minimize the influence of any single rater on the final weights generated.

It is this latter point that is perhaps the most difficult to achieve. There is not at the present time any method that allows the opinions of several experts to be synthesized into a single outcome. In order to accomplish this synthesis, a method for aggregating the inputs of multiple experts is required.

METHODOLOGY

The approach to solving the problem of allowing multiple experts to generate criteria weights was to create a system in which experts could provide their inputs in a symbolic fashion. The symbolic system consisted of listing the criteria from left to right with the most important criteria being in the leftmost position and the least important in the far right position. The following symbols are used to define the relationship of the left criterion to its right neighbor:

= the criteria are of about equal importance
> the left criterion is slightly more important
>> the left criterion is more important
>>> the left criterion is far more important

For a scheme with 12 criteria, the following is an example of how a single evaluator might rank the criteria:

5>6=4>>7>>>1=2>3>>8>>>9=10>11>>13=12

Note that each criterion is used only once and that all criteria are ranked.

Figure 2. shows the list of criteria tentatively identified for evaluation of Halon 1301 replacements. Figure 3. shows how the ratings from 4 evaluators could be "collapsed" in order to produce a single, synthesized, rating. It is the synthesized rating that would then be utilized to generate the weights. The weight generation scheme could then use one of several standard approaches. A simulated decision analysis matrix is shown in Figure 4.

The final stage in the decision system is to perform a sensitivity analysis to determine how the outcome may be affected by variation of the weights. The sensitivity analysis can be conducted using one of several methods, the details of which are unimportant here.

CONCLUSIONS

The decision analysis matrix system described in this report is a new approach to generating weights for a wide variety of applications. The application used to test this methodn is the selection of a Halon 1301 replacement agent. However the main outcome is that the general principles described herein can be used for any cases that would benefit from the inputs of multiple experts in the generation of criteria weights.

CANDIDATE CLEAN AGENT SCREENING ANALYSIS (UNWEIGHTED)

(1 = WORST, 3 = BEST)

MEASURE OF MERIT	HALON 1211		PERFLUOROHEXANE PFC-614		HCFC-123/HALOTRON I	
	FACTOR	SCORE	FACTOR	SCORE	FACTOR	SCORE
150 LB EXTINGUISHER FIREFIGHTING EFFECTIVENESS UL BC POOL FIRE 11.5 FTJ ENGINE NACELLE	240	3	EXPECTED >160	2	EXPECTED >160	2
	3 SEC	3	3 SEC	3	3 SEC	3
TOXICITY < HALON 1211 (4-HR RAT LC 50)	12.0%	2	>30%	3	3%	1
APPLICABILITY TO AIRCRAFT CABIN TOTAL FLOOD USES	LIMITED	2	NO RESTRICTIONS	3	LIMITED	2
COMBUSTION PRODUCTS						
PLUME SEMI-VOLATILES (UG/M3)	219.5	2	58.2	3	411.1	1
HCl/HBr/HF (PPM)	265	2	120.0	3	360	1
SUPERTOXINS (PG/M3)	638	2	68.2	3	961	1
CARBONYL FLOURIDE (PPM) (AVE. CONC. WORST CASE)	14.5	3	96.3	1	39.7	2
FUTURE REGULATORY PRODUCTION PHASE-OUT						
ODP < 0.20	3.0	1	0.0	3	0.02	2
GWP	8-10.0	2	500+ YRS	1	0.02	3
HCFC's	N/A		N/A		POSSIBLE ACCELERATED HCFC PHASE-OUT	
EPA "SNAP" APPROVAL RISK (1990 CAA AMENDMENTS)	NOT POSSIBLE	1	PENDING	3	PENDING	3
CONVERSION COST (14,000 EXTINGUISHERS)	0.0	3	\$0.7M	2	\$0.7M	2
COST PER LB	\$11-27 W/TAX PENALTIES	1	\$10-12	3	HCFC-123 \$4.50 HALOTRON I \$10-12	N/A 3
ACQUISITION CONTRACT FEASIBILITY FOR REPLACEMENT CHEMICAL	NEAT AGENT SINGLE CHEMICAL	3	NEAT AGENT SINGLE CHEMICAL	3	HALOTRON I PROPRIETARY AGENT	1
RAW SCORE		30		36		27

Figure 1.

LIST OF CRITERIA

-  Fire Suppression Efficiency
-  Low Residue Level
-  Low Electrical Conductivity
-  Low Metals Corrosion
-  High Materials Compatibility
-  Stability Under Long Term Storage
-  Low Toxicity
-  ODP
-  GWP
-  Cost
-  Production Availability
-  Extinguishment Concentration
-  Conversion Cost of Facility

Figure 2.

RATINGS INTERPRETATIONS



4 SURVEY SAMPLE

1 >>10 > 7 = 8 = 9 > 12 = 13 > 5 > 11 > 2 = 3 > 4 > 6

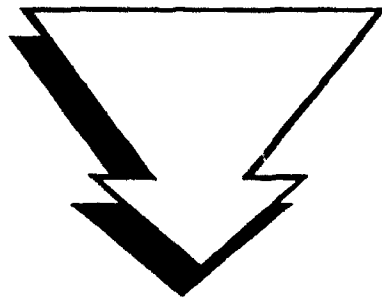
1>>>10>12 > 8 = 7 = 9 > 5 =13 > 2 > 11= 3 > 6 > 4

1 = 10>> 8 = 9 > 7 > 12 = 13 > 5 > 11 > 3 > 2 > 6 > 4

10> 1 >>>7 = 8 > 9 = 12 >>13 = 5 > 6 > 3 > 11> 2 > 4



1 >10 > 7 = 8 = 9 > 12 >13 > 5 > 11 > 2 > 6 > 3 > 4



GENERATE WEIGHTS

Figure 3.

SIMULATED DECISION ANALYSIS MATRIX

	WEIGHT	HALON 1301	C ₆ F ₁₀	FC-218	C-318
COST	4.5	4	4	3	3
F. F. EFFICIENCY	4.8	4	4	3	3
LOW TOXICITY	3.1	4	5	4	4
ODP	2.7	2	4	3	3
SCORE		55.0	63.5	48.4	48.4

Figure 4.

**HYDRAULIC CONDUCTIVITY VARIABILITY,
KRIGING, TREND SURFACES
AND
TRAVELLING WAVES WITH
NONLINEAR, NONEQUILIBRIUM ADSORPTION**

**Valipuram S. Manoranjan
Associate Professor
Department of Pure and Applied Mathematics**

**Washington State University
Pullman, Washington 99164-3113**

**Final Report for:
Summer Research Program
Tyndall Air Force Base Civil Engineering Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

HYDRAULIC CONDUCTIVITY VARIABILITY, KRIGING, TREND SURFACES
AND
TRAVELLING WAVES WITH NONLINEAR, NONEQUILIBRIUM ADSORPTION

Valipuram S. Manoranjan
Associate Professor
Department of Pure and Applied Mathematics
Washington State University

Abstract

The spatial variability of hydraulic conductivity is studied using the (grain-size analysis) data collected at a tracer test site, located at Columbus Air Force Base in Mississippi. The study involves vertical kriging and constructing segmented planar trend surfaces. It is shown that such an investigation can provide a better understanding of the variability pattern of hydraulic conductivity. Further, a relationship between the results from borehole flowmeter study and grain-size analysis study is conjectured. Lastly, a one dimensional transport model with nonlinear, nonequilibrium adsorption is analyzed and an expression for the travelling wave solution is obtained in a closed form employing a linear caricature approximation of the nonlinear isotherm.

HYDRAULIC CONDUCTIVITY VARIABILITY, KRIGING, TREND SURFACES AND TRAVELLING WAVES WITH NONLINEAR, NONEQUILIBRIUM ADSORPTION

Valipuram S. Manoranjan

1. INTRODUCTION

It is generally recognized that the heterogeneity of the aquifer hydraulic conductivity plays an important role in controlling the dispersion and movement of groundwater solutes. So, there is a great need to quantify spatial variability of hydraulic conductivity if existing transport models are to be applied to practical problems. However, most of the time one would not have a large number of hydraulic conductivity measurements, and the study of quantifying variation in hydraulic conductivity will be based on a limited data set. The main objective of this project is to carry out such a variability study on the data collected at a tracer test site, located at Columbus Air Force Base in Mississippi. In the recent past, somewhat similar studies have been conducted by other researchers [Rehfeldt et al., 1989 & 1991; Young et al., 1992]. These studies were essentially based on borehole flowmeter measurements of hydraulic conductivity, and no serious attempt was made to carry out a thorough study involving indirect measurements such as soil grain-size analyses. We believe that our study is the first serious attempt in using hydraulic conductivity estimates obtained from grain-size analyses to understand spatial variation in hydraulic conductivity. Our work will also look for some kind of relationship between the results one will obtain from borehole flowmeter study and grain-size analysis study.

The standard approach of tackling a problem of the kind proposed here is to introduce the concept of regionalized variable ($\ln[\text{hydraulic conductivity}]$ in our case), and to do kriging, a local estimation technique which provides the best linear unbiased estimator, on the regionalized variable over the site (a three dimensional domain) of interest. We feel, sometimes, such a three dimensional kriging study might obscure any simple behavioral patterns the regionalized variable might have in one particular direction or the other. Therefore, we take

a fresh look at this approach introducing segmentation in the data. Still, using kriging, but in one direction only, and employing the concept of trend surfaces, we find that spatial variation in hydraulic conductivity can be better understood with a limited data set.

As we pointed out earlier, quantification of spatial variability of hydraulic conductivity is essential for employing transport models to practical problems. However, one should not naively think that once spatial variation pattern of hydraulic conductivity is available, applying transport models is an easy task. There are various unanswered questions associated with transport models and the one we are particularly interested in asks, how to describe transport, analytically, with nonlinear and nonequilibrium adsorption. As a secondary objective of the summer project we have tried to answer this question. By studying the one dimensional transport in the positive z direction (along the depth), and using an appropriate linear caricature to describe the nonlinear adsorption, we are able to find the closed form solution for the concentration profile of the solute. It does not matter whether the adsorption is *Langmuir* or *Freundlich*, the obtained solution can be easily modified to suit our needs. Such a solution will be very useful for predicting the concentration of a solute in any location at any given time without resorting to expensive numerical computations.

2. ANALYSIS OF EXPERIMENTAL DATA

Hydraulic Conductivity Estimates

The experimental data are from soil samples obtained from 17 separate locations at the Columbus Air Force Base tracer test site. At each location, samples are collected at 9 different (irregularly spaced) vertical depths. An empirical formula which relates hydraulic conductivity to grain size is used to estimate conductivity values for these 153 soil samples. We use the formula in the form presented in [Seiler, 1973]. The formula is of the form,

$$K = fd^2$$

where, K is the hydraulic conductivity, d , the representative grain diameter and f , the

proportionality factor. The factor f is a function of the uniformity coefficient U given as,

$$U = d(60)/d(10)$$

where, $d(m)$ is the grain diameter such that $m\%$ of the sample by weight is of diameter less than $d(m)$. Following [Seiler, 1973], when $U \geq 5$, our hydraulic conductivity estimates are based on the formulae

$$K = f(U) d(10)^2, \quad 5 \leq U \leq 17$$

and

$$K = f(U) d(25)^2, \quad U > 17.$$

The values for $f(U)$ can be obtained from either [Seiler, 1973] or [Rehfeldt et al., 1989]. However, when $U < 5$, we use the formula

$$K = d(10)^2$$

due to [Hazen, 1892].

Punctual Kriging Over Depth

In order to analyze the spatial variability of the hydraulic conductivity estimates thus obtained using the empirical formulae, we introduce the concept of regionalized variable. A regionalized variable can be thought of as an intermediate variable between a truly random variable and a completely deterministic one. Unlike random variables, a regionalized variable will have continuity from point to point, but usually it will not be possible to know its value everywhere, as one could for a deterministic variable. Here, the obvious candidate for the regionalized variable is the hydraulic conductivity K . However, geostatistics requires that the regionalized variable be normally distributed, and our hydraulic conductivity estimates are not normally distributed. Therefore, a better choice for the regionalized variable is $\ln(K)$. Usual practice is to assume stationarity of the regionalized variable and to find the best estimator (with minimum variance) of the mean value of the regionalized variable (punctual kriging) over a two or a three dimensional domain. Thus obtaining some variability pattern of the regionalized variable. The currently available software packages such

as GEOPACK and SURFER employ this approach too. The problem with this approach is that, even though, one will end up with some kind of resulting variability pattern of the regionalized variable after a lot of computations, interpreting this pattern may not be easy. This pattern may not fall into the category of known simple patterns that can be described by polynomial functions, sinusoidal functions etc. So, making any predictions based on a non-understandable pattern may not seem a good idea.

We take a completely different approach in analyzing the hydraulic conductivity data. We would like to investigate whether there exists any known simple patterns of either K or $\ln(K)$ in a given direction. Therefore, to start with we will not carry out any two or three dimensional kriging, but, just a one dimensional kriging over depth at each location of interest. The first step in the kriging process involves calculating the semi-variogram associated with the given experimental data. At this point, the basic assumption is that the log hydraulic conductivity field is stationary. The semi-variogram Y will describe the expected difference in value between pairs of samples with a given relative orientation. If we denote the regionalized variable $\ln(K)$ by R , the formula that is used to calculate the semi-variogram is given by

$$Y(h) = \frac{1}{2n} \sum_{i=1}^n [R(z_i) - R(z_i + h)]^2$$

where, z_i and $z_i + h$ are two different depths at a distance h apart and n is the number of sample pairs separated by this distance h , known as the lag distance. The graph Y vs. h is the semi-variogram. The computed Y values can be fitted by various models and in most studies linear models are used (as in the software package SURFER) just for the sake of convenience.

In this preliminary study, we carry out kriging at three different locations and it turns out that the corresponding semi-variograms are nonlinear with the nugget effect (i.e. discontinuity at the origin). A typical semi-variogram is shown in Figure [1]. Once the form of the semi-variogram is decided, the next step is to do kriging itself. By doing kriging, what we obtain is the estimated value for the regionalized variable R at any depth in a given

location. For a detail account of kriging, the interested reader should consult [Journel & Huijbregts, 1991]. At a given location, if the number of points employed for kriging is m , then the estimate value at any point p will be given by

$$R_p = W_1 R_1 + W_2 R_2 + \dots + W_m R_m.$$

The weights W_i 's are not known at this stage and they will be evaluated by solving an appropriate system of linear equations,

$$AU = b.$$

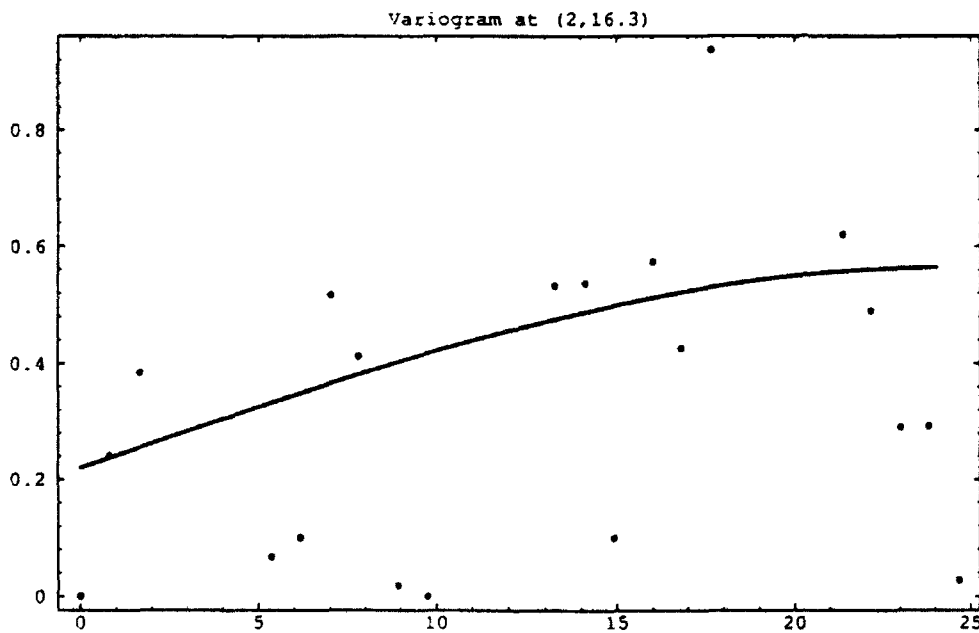


Figure 1

The elements in the matrix A , and the vector b are calculated using the semi-variogram and the unknown vector U is composed of the weights W_i 's and the Lagrange multiplier λ . The interesting aspect of kriging is that, if we choose the point p to be any one of the known points 1 to m , the estimate we obtain will be exactly the same as the given value. (i.e. for example if $p = 3$, the kriging process will give the weights as $[W_1, W_2, W_3, W_4, \dots, W_m] = [0, 0, 1, 0, \dots, 0]$, and thus the estimate is obtained as R_3 exactly). In our study, by doing some simple mathematics we are able to prove the correctness of this nice kriging characteristic.

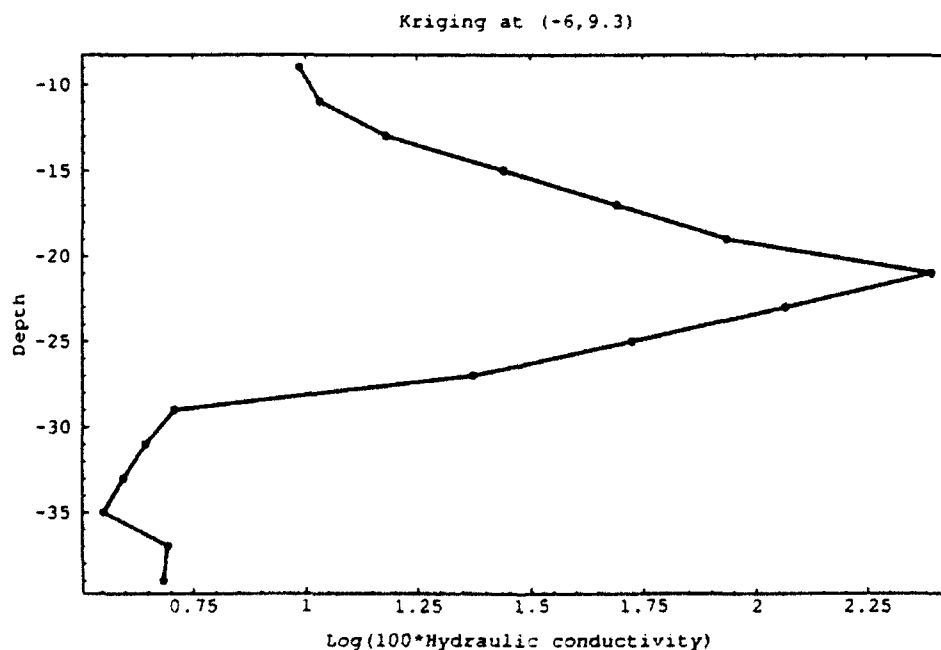


Figure 2

At each of the chosen three locations, we carried out kriging at equally spaced depths and the final results are presented in figures [2], [3] and [4]. It is interesting to note that the qualitative behavior seems to be the same in all these figures. (i.e. with the increase in depth, hydraulic conductivity increases for a while, then sharply decreases and slowly increases again). In figure [5] we present all three graphs in one frame. Now, comparing the curves corresponding to the locations (2, 16.3) and (7, 14.6), we may conclude that the variation in the hydraulic conductivity becomes smaller as the x value increases for a nearly constant y value. However, we hasten to add that it is necessary to carry out further work on this aspect before accepting this tentative conclusion.

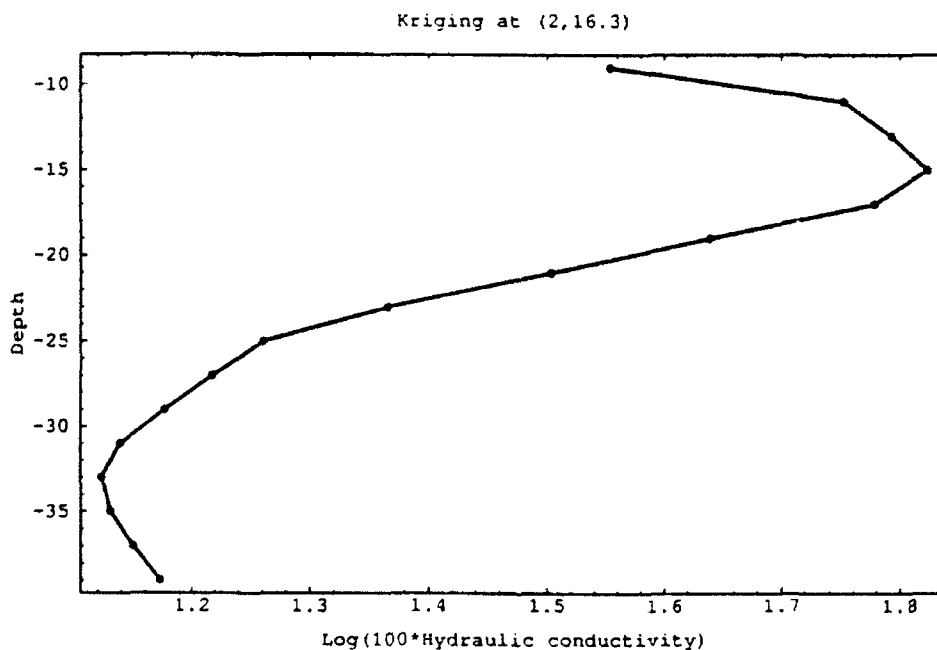


Figure 3

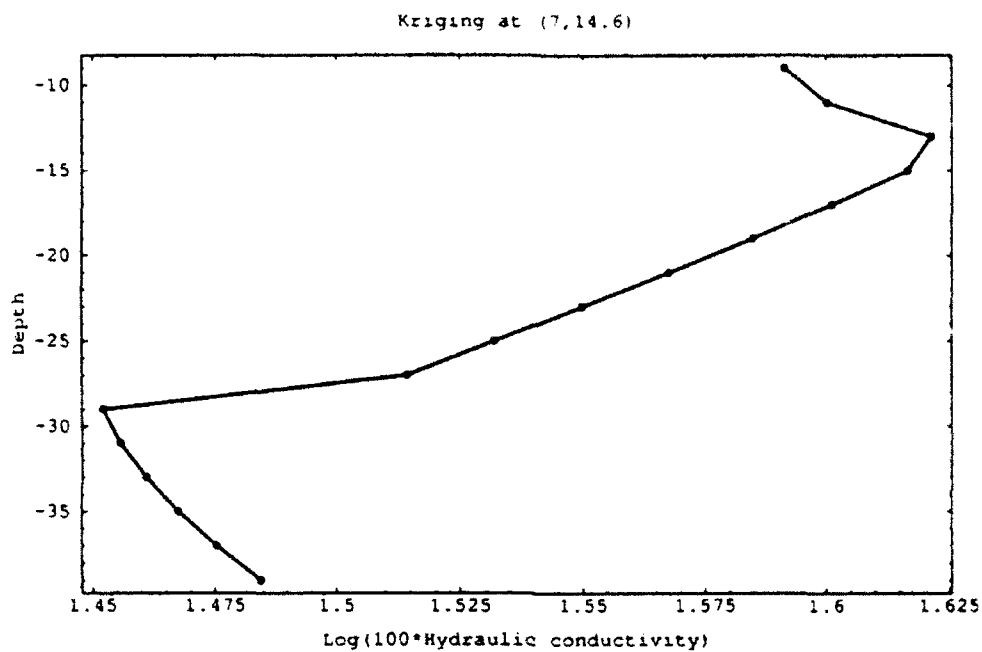


Figure 4

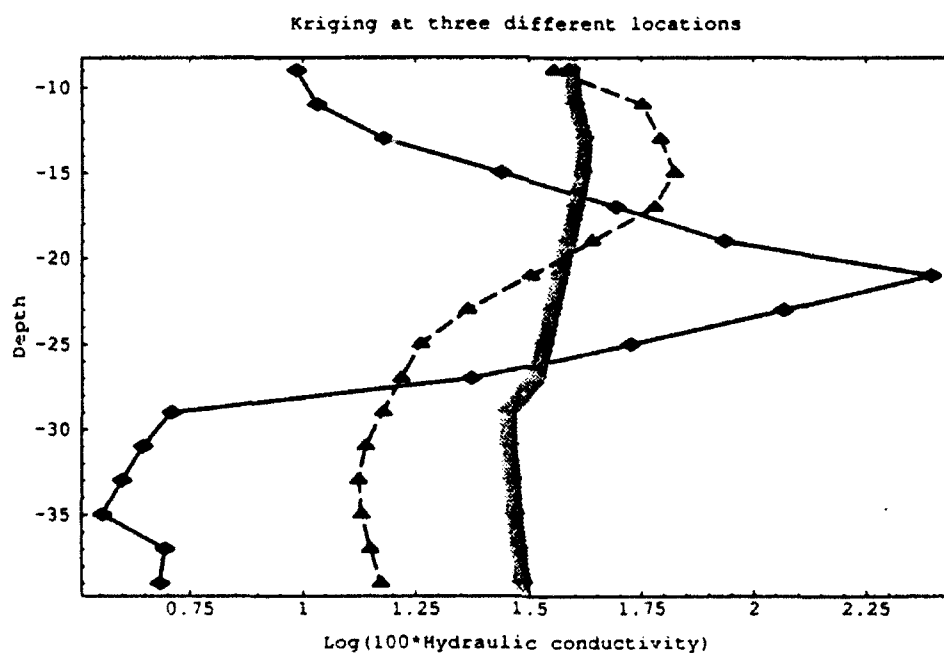


Figure 5

Trend Surfaces

The way we have applied kriging gives us only the variability pattern of hydraulic conductivity in the z direction (depth). In order to investigate the spatial pattern in the x, y directions, we divide the tracer test site into three segments. We call these segments nearfield, midfield and farfield depending on how close or how far they are from the point of tracer injection. Now, picking one of these field segments, at a chosen depth we can construct a trend surface for hydraulic conductivity based on the sample data collected just in that field segment. In figures [6], [7] & [8] we present such typical trend surfaces obtained for nearfield, midfield and farfield, respectively. Initially, linear trend surfaces were tried, but the percentages of goodness of fit were not good enough. However, for second order trend surfaces, in all three segments, the coefficients of determination turn out to be approximately equal to 0.99. Therefore, we feel that the second order trend surfaces are the best possible low order trend surfaces for the experimental data we have. One might ask, what happens if a trend surface is constructed to the whole tracer test site without segmenting it? In order to answer this, we tried a second order trend surface for the whole site, at a depth of 28.5 feet, and the corresponding value of the coefficient of determination was about 0.44, which is not a very good value. Then we tried a third order trend surface and obtained the value 0.80 for the coefficient of determination. Although 0.80 is a better value than 0.44, it is still not that good compared to 0.99 obtained for segmented trend surfaces. One can always go another order higher and hopefully find a better trend surface for the whole site. However, it is not advisable to go to higher and higher orders, because that will increase the complexity of the problem. Further, when one has only a limited number of data points, fitting higher order trend surfaces may introduce unwanted artifacts. Therefore, it is our contention that segmenting the test site and fitting a trend surface in each of the different segments is how the $x - y$ spatial variability of hydraulic conductivity should be studied.

This approach we have taken constructs second order trend surfaces in the $x - y$ plane and kriges in the z direction. This means that we are able to see nice second order polynomial patterns of hydraulic conductivity in the $x - y$ plane, and a simple looking variation of

K in the z direction. Based on the limited analytical study we carried out this summer, our approach seems a simple and reasonable way of characterizing the spatial variation of hydraulic conductivity. However, we should point out again the necessity to carry out more work before making any firm conclusions. Also, one should remember that kriging is a local estimation procedure, and the kriged estimates are valid only in the neighborhood of the data points considered. Similarly, the trend surfaces cannot be extended to regions far away from the points employed in these surface constructions.

Before we close this section, a couple of important points should be made. The first relates to the grain-size data we have used in our study. One may wonder how good the results obtained from these data are compared to the results from borehole flowmeter data. In fact, in a recent paper [Rehfeldt et al., 1991] on hydraulic conductivity variability, the argument was made in favor of borehole flowmeter data. The argument was based on the fact that the estimates of log hydraulic conductivity variance and the horizontal & vertical correlation scales obtained for grain-size data did not fall within the confidence limits of the estimates from borehole flowmeter data. For borehole flowmeter data the respective estimates were found to be 4.5, 12m and 1.5m; whereas, for grain-size data these values were 2.9, 15m and 0.8m. These estimates were found assuming second order stationarity of conductivity field. However, in the same paper at a later section, the authors argue about the need to introduce trends in hydraulic conductivity analysis, because of the observed large-scale spatial variations in the mean ground water velocity. This will invalidate the earlier stationarity assumption for conductivity field, but now log conductivity residue field will be assumed to be stationary. Analyzing further, the authors introduce a third order polynomial trend to represent conductivity drift based on its compatibility with groundwater flow field. Under these circumstances, working with borehole flowmeter data, the values they obtain for variance and the horizontal & vertical correlation scales (for log conductivity residuals) are respectively 2.8, 5.3m and 0.7m. The values 2.8 and 0.7m for variance and vertical correlation scale are almost the same as the respective values 2.9 and 0.8m they obtained using grain-size data! Also, the value 5.3m for the horizontal correlation scale is nearly the same

as the value one will obtain when (grain-size) horizontal correlation scale 15m is divided by 3, the order of the polynomial trend!

This clearly demonstrates that there exists an interesting relationship between the results obtained from grain-size data and the de-trended flowmeter data. However, further extensive study is needed to make any concrete statements on this. But, at least this simple argument shows that one should not discard grain-size data as useless for the study of hydraulic conductivity variability.

The second point we would like to make is that, although our work in this summer project has employed only the grain-size data, the approach we have taken (segmented trend surfaces in the $x - y$ plane and vertical kriging) is equally applicable to any other hydraulic conductivity data that can be generated using either direct or indirect methods.

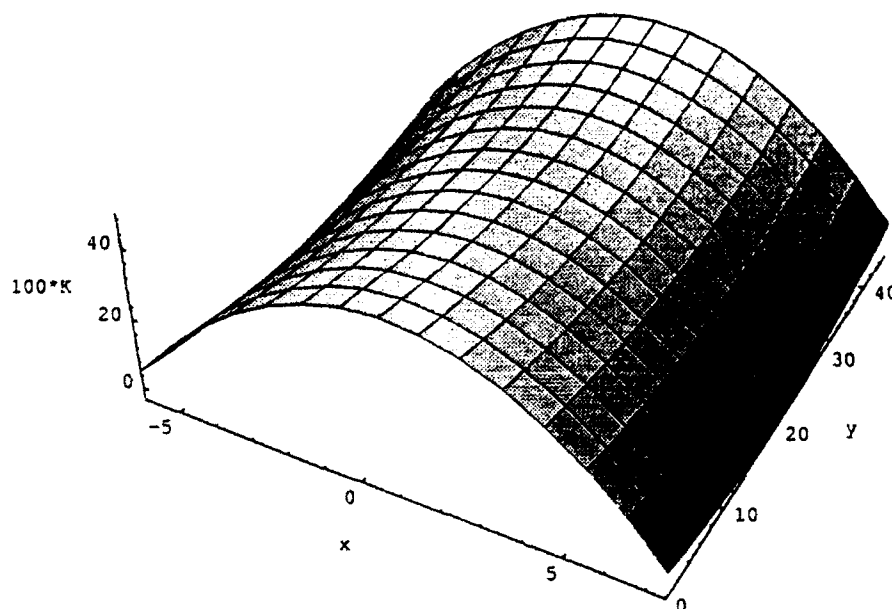


Figure 6
Near Field Trend Surface for Hydraulic Conductivity (Second Order; $R^2 = 0.999$)

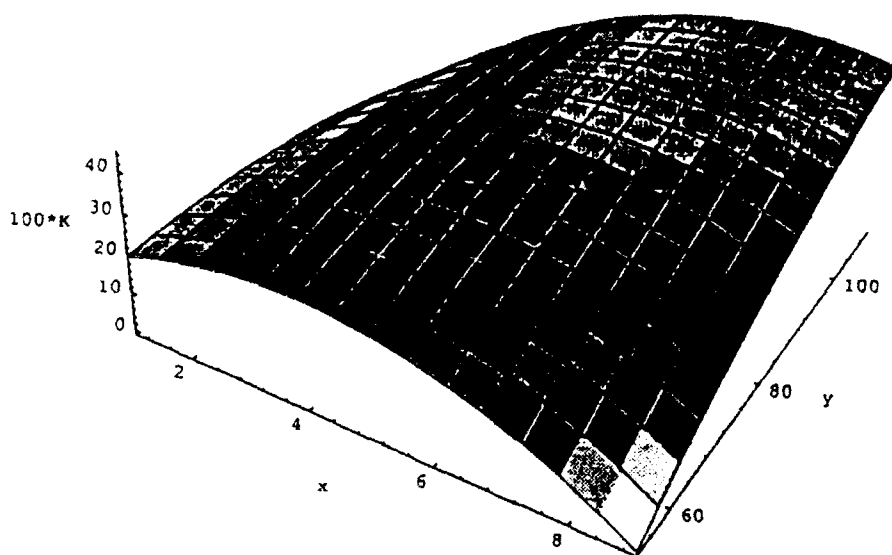


Figure 7
Mid Field Trend Surface for Hydraulic Conductivity (Second Order; $R^2 = 0.999$)

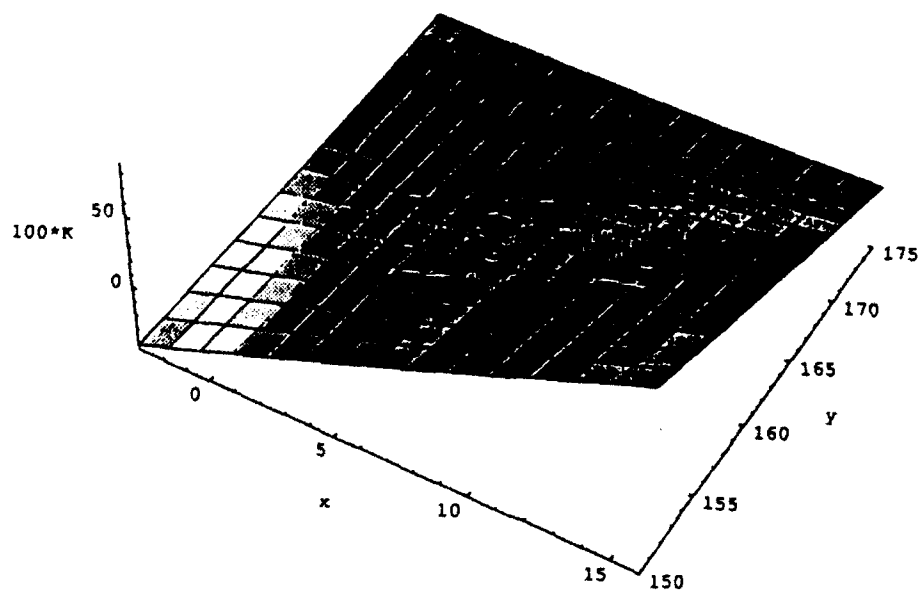


Figure 8
Far Field Trend Surface for Hydraulic Conductivity (Second Order; $R^2 = 0.999$)

3. CONCENTRATION PROFILE/NONLINEAR ADSORPTION

The second part of our work was on a one dimensional transport model. The investigation was motivated by the need for analytical expressions of solute concentration profiles in order to make long term predictions with minimum computational cost and effort. In the past, various authors [van der Zee, 1990 and references therein] have studied transport models with the aim of obtaining analytical expressions for concentration profiles. Their efforts have been only partially successful, because analytical expressions were found possible only under local equilibrium assumption or when there was linear sorption.

In our study, we analyse a solute transport model with nonlinear sorption, when sorption kinetics is described by a first order rate expression. In a non-dimensionalized form, the one dimensional model in the positive z direction (depth) can be described as,

$$\begin{aligned} \text{[Rate of change in concentration]} &= \text{[(Dispersion)/(Peclet number)]} - \text{[Convection]} \\ &- \text{[(a parameter) * (Rate of change in adsorbed amount)]} \end{aligned}$$

with

$$\text{[Rate of change in adsorbed amount]} = \text{[a rate constant] * [Nonlinear sorption - Adsorbed amount]}.$$

The nonlinear sorption in the above model can be Langmuir, given as,

$$KQC/(1 + KC)$$

or Freundlich of the form,

$$KC^n,$$

where C is the solute concentration, K , the nonlinear adsorption coefficient, Q , the Langmuir adsorption maximum and n , Freundlich sorption exponent.

We investigate the possibility of constructing travelling wave solutions for solute concentration C . In order to do that, first, we replace the nonlinear sorption function in the model

by a suitable linear caricature which has the same qualitative behavior. In particular, we use the linear caricature of the form,

$$f(C) = \begin{cases} K_1 C, & C \leq C^* \\ A + K_2 C, & C \geq C^* \end{cases}$$

with appropriate values for the constants A , K_1 and K_2 . It is obvious that when we choose $K_2 = 0$ with suitable values for A & K_1 , $f(C)$ will have the Langmuir behavior. The concentration value C^* is chosen based on experimental evidence.

Our study involves introducing the travelling wave coordinate

$$Z = z - vt$$

and solving the resulting third order differential equations in the segments $(-\infty, 0]$ and $[0, \infty)$ respectively. By using the matching conditions at the origin, we are able to obtain the continuous concentration profile we were seeking. For the parameter values of [van der Zee, 1990], the concentration profile is as shown in Figure [9]. The corresponding linear caricature is given in Figure [10].

It should be noted that one can further improve the choice of the linear caricature. For example, by constructing one with three line segments instead of two. Our mathematical treatment can still be applied in such a situation, but, with some extra amount of work. The mathematical ideas employed in this part of our study will be very useful for long term predictions of solute transport over large distances or times. We should point out that comparative studies using experimental data are needed for further validation of the results obtained here.

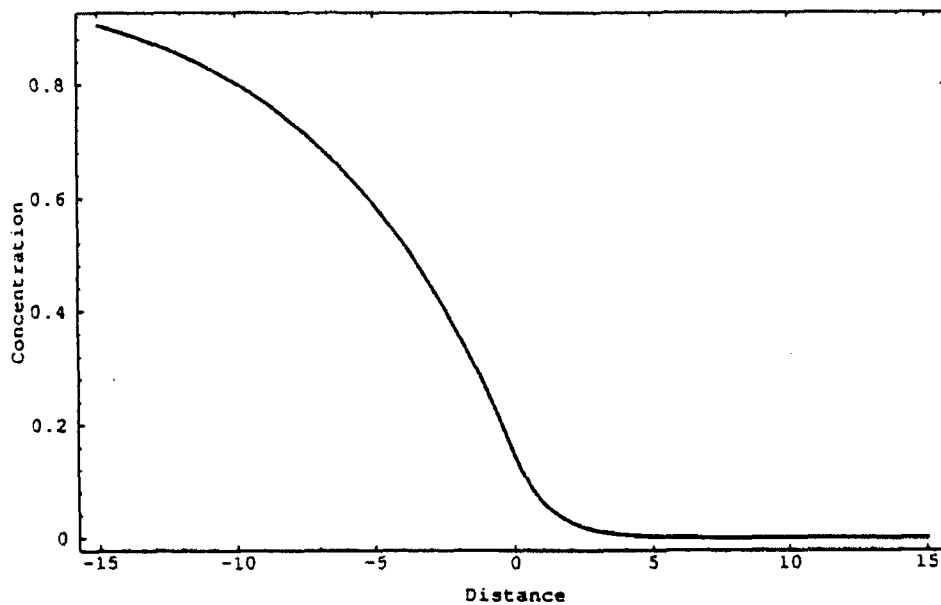


Figure 9
 'Freundlich Wave' from Linear Caricature

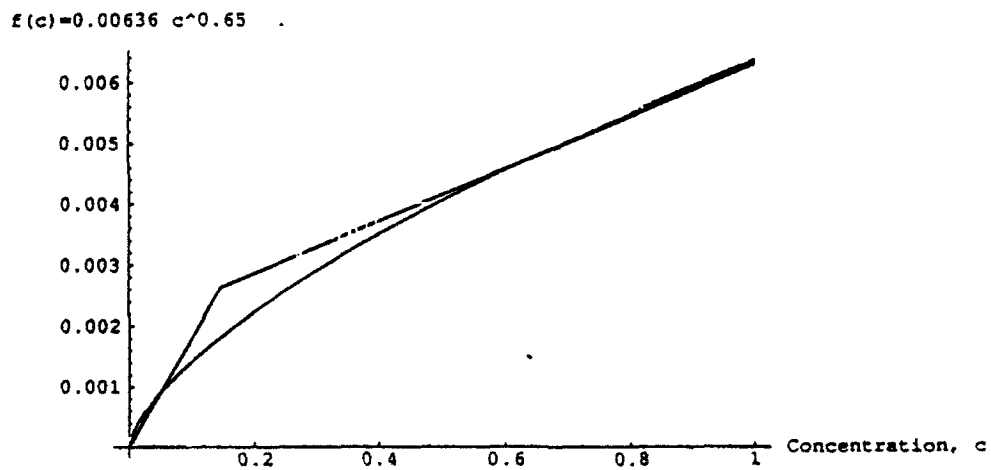


Figure 10
 Freundlich isotherm & a linear caricature

4. CONCLUSIONS

Although, the hydraulic conductivity variability study conducted here is just a limited, preliminary study, we are able to make the following two important conclusions:

- (i) The idea of vertical kriging and segmented trend surfaces can provide a simple and understandable view of hydraulic conductivity variability pattern.
- (ii) The grain size analysis data should not be discarded as useless in the investigation of hydraulic conductivity variability, since the use of grain-size analysis data in such a variability study can give results which can be related to the results one obtains from the (supposedly more accurate) borehole flowmeter data study.

In the work on the transport model with nonlinear, nonequilibrium adsorption, a closed form solution is obtained for the concentration wave. An expression of this nature will be very useful for predictive purposes and will help experimenters avoid costly numerical computing associated with nonlinear isotherm problems.

5. FUTURE WORK

As noted in the section on punctual kriging, when a limited study of kriging was carried out over depth in the near field, it was observed that the variation in hydraulic conductivity became smaller with an increase in x -value when y -value was held nearly constant. In order to decide whether this is a global behavior or not, one needs to carry out further kriging in mid and far field locations.

In the section on trend surfaces, based on a deductive argument we made a conjecture relating the results from borehole flow meter study and grain-size analysis study. Again, this conjecture can be justified only if a thorough study of borehole flowmeter data is conducted in conjunction with our grain-size analysis data and compared.

One of the major problems with the kind of kriging-trend surface analysis carried out in this work is that our variable (hydraulic conductivity) may not have been sampled sufficiently to provide estimates of acceptable precision. In such circumstances, the estimation may be improved by considering spatial correlations between our variable and some other sample variable. One may be able to perform such a co-kriging analysis combining both the hydraulic conductivity data and TOC (toxic organic content) data.

Finally, when obtaining a model for the variogram, maybe ideas such as hole effects and nested-variograms could be incorporated to get a better fitted variogram model.

REFERENCES

1. Rehfeldt, K.R., L.W. Gelhar, J.B. Southard, and A.M. Dasinger, Estimates of Macrodispersivity based on analyses of hydraulic conductivity variability at the MADE site, *EPRI Report EN-6405, Research Project 2485-5*, 1989.
2. Rehfeldt, K.R., J.M. Boggs and L.W. Gelhar, Field study of dispersion in a heterogeneous aquifer: 3. Hydraulic conductivity variability, Preprint, 1991.
3. Young, S.C., J. Herweijer and D.J. Benton, Geostatistical evaluation of a three-dimensional hydraulic conductivity field in an alluvial terrace aquifer, Preprint, 1992.
4. Seiler, K.P., Durchlässigkeit, Porosität und Kornverteilung quartärer Keis-Sand- Ablagerungen des bayerischen Alpenvorlandes, *Gas-und Wasserfach*, **114(8)**, 1973, 353-400.
5. Hazen, A., Experiments upon the purification of sewage and water at the Lawrence experiment station, *Mass. State Board of Health, 23rd Annual Report*, 1892.
6. Journel, A.G., and Ch. J. Huijbregts, Mining Geostatistics, Academic Press, 1991.
7. Van der Zee, S.E.A.T.M., Analytical Traveling wave solutions for transport with non-linear & nonequilibrium adsorption, *Water Resources Research*, **26**, 1990, 2563-2578.

**IMMOBILIZED CELL BIOREACTOR FOR
2,4-DINITROTOLUENE DEGRADATION**

Kenneth F. Reardon
Assistant Professor
Department of Agricultural and Chemical Engineering

Colorado State University
Fort Collins, CO 80523

Final Report for:
AFOSR Summer Research Program
HQ AFCESA/RAVC

Sponsored by:
Air Force Office of Scientific Research
Tyndall Air Force Base, Panama City, FL

September 1992

IMMOBILIZED CELL BIOREACTOR FOR 2,4-DINITROTOLUENE DEGRADATION

Kenneth F. Reardon
Assistant Professor
Department of Agricultural and Chemical Engineering
Colorado State University

Abstract

2,4-Dinitrotoluene (DNT) is an EPA priority pollutant that is found in wastes from the manufacture of 2,4,6-trinitrotoluene (TNT) and toluenediisocyanate. Treatment of DNT-containing waste by activated sludge processes is usually difficult owing to the high toxicity of this compound. In this work, removal of DNT from aqueous solutions was achieved with a bioreactor containing immobilized *Pseudomonas* PR7. This microorganism, which is able to use DNT as its sole source of carbon and energy and degrades DNT oxidatively, was isolated by the Microbiology Group of HQ AFCEA at Tyndall AFB.

Studies using suspended cells showed that *Pseudomonas* PR7 is able to degrade DNT concentrations as high as 1500 μM , the solubility limit at 25 °C, although concentrations above 1000 μM lengthened the lag phase in batch growth. Higher cell densities were obtained on a growth medium containing both succinate and DNT than when DNT was the sole carbon source, but 12-fold higher specific rates of DNT degradation were obtained in succinate-free cultures. High concentrations of nitrite (>5000 μM) inhibited the growth of this microorganism but did not affect the specific DNT degradation rate.

The packed-bed bioreactor contained diatomaceous earth pellets to which the microorganisms were attached. The feed to the system contained DNT (from 500 to 1200 μM) as the sole carbon source in a minimal medium containing phosphate, ammonium, and other nutrients. During a 712-hour test of this bioreactor, various studies were carried out to characterize the system. When the system was configured to emulate a continuous well mixed reactor, the maximum performance was 325 $\mu\text{mol/L}\cdot\text{h}$ (based on total liquid volume) or 470 $\mu\text{mol/L}\cdot\text{h}$ (based on the column volume). The comparable rate in a continuous suspended cell reactor was 25 $\mu\text{mol/L}\cdot\text{h}$. The bioreactor also demonstrated resistance to large perturbations and the ability to function stably for long periods. When fed a mixture of DNT and TNT, both were degraded, and the rate of DNT consumption was unaffected by the presence of TNT.

IMMOBILIZED CELL BIOREACTOR FOR 2,4-DINITROTOLUENE DEGRADATION

Kenneth F. Reardon

INTRODUCTION

A. Bioreactors for Pollution Control

Many hazardous organic pollutants can be degraded to innocuous compounds by microorganisms [1]. This concept forms the basis for the set of technologies called "bioremediation", of which two major types can be identified: in situ bioremediation, in which the contaminated soil and water is treated without being moved, and bioreactor treatment, in which microorganisms contained in a reactor are used to treat flows of contaminated soil and/or water. For the cleanup of large areas or deep aquifers, in situ bioremediation approaches are often better; however, soil-wash streams, industrial waste streams, and some aquifers are more appropriately treated in bioreactors.

Bioreactors are divided into those containing either suspended (free) cells or immobilized cells [2]. Suspended cell reactors, by far the more common, can be operated in batch (including fed- and sequencing batch) or continuous flow modes (Figure 1). For biotreatment of waste streams, the former has the advantage of simplicity of operation but disadvantages stemming from the time required to fill and drain the reactor vessel and, since the concentration of pollutant is relatively high at the start of each run, lowered activity due to substrate inhibition (a common phenomenon for most molecules of interest in waste treatment). On the other hand, continuous-flow reactors (generally continuous stirred-tank reactors), require no fill and drain time and can be designed to avoid substrate inhibition problems. However, the operation of these systems can be more complex and is always limited by the specific growth rate of the microorganisms serving as the biocatalyst. That is, since the governing equation for a steady-state continuous stirred tank reactor (CSTR) is $\mu = D$, where μ is the specific growth rate and D is the dilution rate (flow rate divided by volume), the maximum flow rate for a given system is dictated by the rate at which the cells are able to grow; excessive flow rates lead to washout of cells from the reactor vessel. Microbial growth rates on many organic pollutants are slow and thus this constraint is a significant one. A further disadvantage of most suspended cell reactor systems used in wastewater treatment is that the low concentrations of carbon sources (pollutants) in the feed

streams can support only small populations of cells, resulting in a low volumetric reactor productivity (amount compound degraded/reactor volume/time).

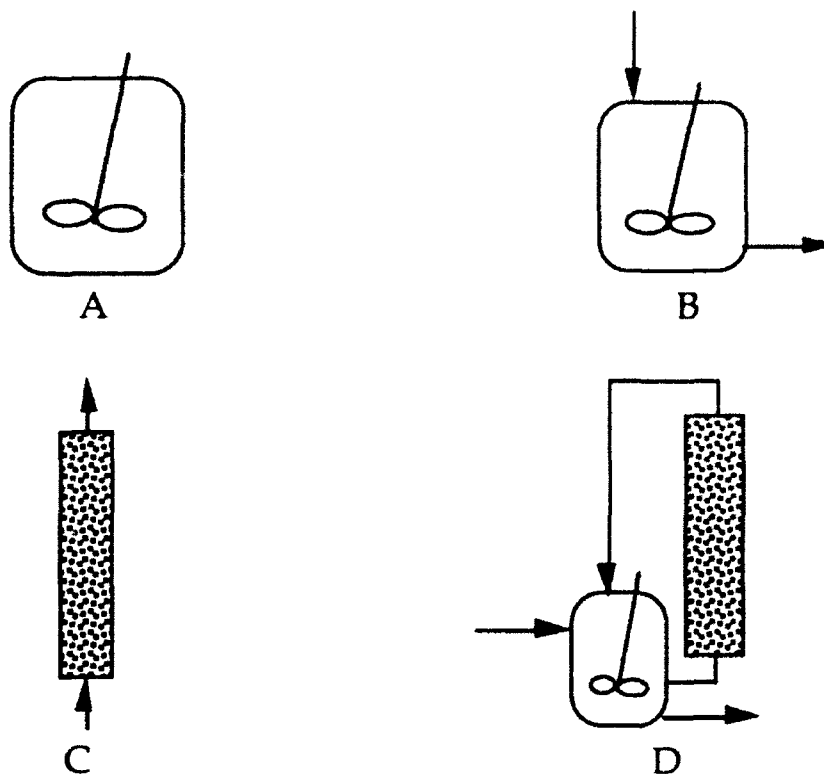


Figure 1. Bioreactor types. (A) suspended cells, batch stirred tank; (B) suspended cells, continuous stirred tank (CSTR); (C) immobilized cells, plug flow (PFR); (D) immobilized cells, CSTR.

Many of the problems with suspended cell bioreactors can be overcome with the use of immobilized cells, which are defined as cells that are confined within a reactor by attachment to the surface of small particles (e.g., sand), entrapment within a porous polymer (e.g., calcium alginate), flocculation, or retention behind a membrane [3]. Bioreactors based on any of these techniques contain high cell densities and thus provide a higher volumetric productivity than comparable suspended cell systems. In addition, the optimization of continuous-flow immobilized cell reactors is not restricted by the specific growth rate of the cells (since they are retained within the reactor), and the effluent from one of these reactors contains a very low concentration of cells.

Immobilized cell bioreactors can be run in a batch mode but are usually used in a continuous-flow configuration. A common example of this type of bioreactor is the packed column, in which pellets of adsorbed or entrapped cells are loaded into a long, narrow tube. The liquid feed may be conditioned (e.g., aeration, pH adjustment) before flowing upward through the cell bed. Continuous-flow packed bed reactors are frequently operated in a single-pass mode that approximates a plug-flow reactor (PFR) system (Figure 1C). Alternatively, the packed-bed reactor can be operated in a recycle mode that approximates a CSTR system (Figure 1D). While the PFR system is simpler and can result in an effluent containing no contaminant, it is subject to substrate inhibition or toxicity problems since the substrate concentration is high at the column inlet. In the packed-bed CSTR system, this substrate inhibition problem is avoided because the feed is diluted when it is mixed with the liquid content of the system. However, a zero concentration of contaminant in the effluent from a CSTR is not possible since this stream has the same composition as the liquid recirculating in the bioreactor. It can be shown that a reactor system consisting of a CSTR (in which the contaminant concentration is reduced from a high to a relatively low level) followed by a PFR (in which the contaminant concentration can be reduced to a very low or zero concentration) results in the highest overall volumetric productivity [4].

B. 2,4-dinitrotoluene (DNT)

2,4-Dinitrotoluene (DNT) is an EPA priority pollutant that is found in wastes from the manufacture of 2,4,6-trinitrotoluene (TNT) and toluenediisocyanate. Treatment of DNT-containing waste by activated sludge processes is usually difficult because of the high toxicity of this compound to most microorganisms and the fluctuations in the DNT concentration entering the treatment plant. Periods of high inlet DNT levels inhibit the growth of the DNT-degrading cells in the reactor and a significant proportion of them may be washed out, lowering the ability of the system to remove DNT. A similar problem occurs when periods of low inlet concentration (during which fewer DNT-degrading cells can be maintained in the reactor) are followed by periods of high DNT feed levels (when the concentration of DNT-degraders is too low to remove DNT to the desired concentration).

DNT has a molecular weight of 182.13, and thus a 500 μM solution is equivalent to 93.9 ppm. The maximum solubility of DNT in water is 1438 μM at 22 $^{\circ}\text{C}$ [5].

PROJECT GOALS

The goals of this project were to develop and test an immobilized cell bioreactor capable of high rates of DNT degradation. To attain these goals, several specific tasks were performed:

- ◆ microorganism selection
- ◆ investigation of the biodegradation kinetics
- ◆ selection of the immobilization method
- ◆ design and construction of the bioreactor system
- ◆ characterization of the bioreactor system, including

MATERIALS AND METHODS

A. Media

The microorganisms used in this work were grown in solutions of DNT in Spain's Mineral Salts Base (SMSB), which contains (mg/L): $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 112.5; $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 5.0; $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 2.5; KH_2PO_4 , 340; $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 670; CaCl_2 , 13.8; FeCl_3 , 125; NH_4Cl , 500. This medium was occasionally supplemented with 10 mM sodium succinate. Strains were maintained on plates containing SMSB, 500 μM DNT, 10 mM succinate, and 18 g/L Bitek agar (Difco) or on plates made from tryptic soy broth (Difco) (3.75 g/L), 500 μM DNT, and 18 g/L Bitek agar.

B. Analytical Methods

The concentrations of DNT, its degradation intermediate 4-methyl 5-nitrocatechol, and TNT were determined by high pressure liquid chromatography (HPLC). Separations were achieved on a phenyl column (Bio-Rad) with a 1.5 mL/min gradient of acetonitrile and trifluoroacetic acid (13.5 mM) as the mobile phase. The elution gradient began with 40% acetonitrile/60% aqueous trifluoroacetic acid and was changed to 60% acetonitrile over 3 min, where it was held for an additional 5 min. Compound detection (230 nm) and identification were performed using a diode array detector. The injection volume was 100 μL .

Nitrate concentrations were measured using the sulfanilic acid/sulfanilamide method [6]. Succinate concentrations were determined by ion chromatography. Cell densities were estimated by measuring the optical density at 600 nm.

RESULTS AND DISCUSSION

A. *Microorganism Selection*

Several DNT-degrading strains have been isolated by the Microbiology Group of HQ AFCEA. Three of those, *Pseudomonas* DNT-1, *Pseudomonas* PR7, and *Pseudomonas* R34, were screened for use in the bioreactor. All three strains can utilize DNT as their sole source of carbon and energy via an oxidative metabolic pathway that has been partially elucidated (Figure 2) [7]. As can be seen, the two nitro groups are eliminated as nitrite, which can be used to verify the amount of DNT that has been degraded [7]. Although there appear to be some differences among these strains in the regulation of this pathway, the necessary enzymes in all three can become uninduced in the absence of DNT. When DNT is reintroduced, the sequential induction of enzymes results in the temporary accumulation of 4-methyl 5-nitrocatechol (MNC), a bright yellow compound. Transient accumulation of MNC was also observed when the concentration of DNT was increased rapidly. The length of time required for reinduction of the complete pathway could be problematic for treatment of fluctuating feeds.

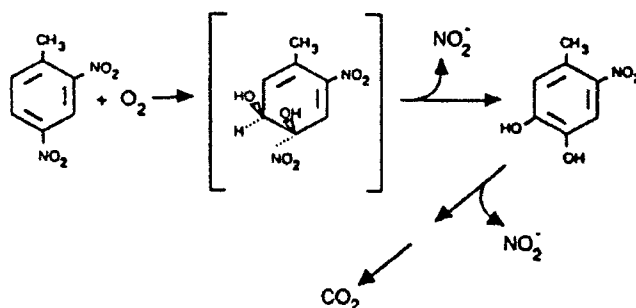


Figure 2. Oxidative metabolic pathway for DNT degradation [7].

In order to compare their activities, the three strains were grown at 32 °C in flask cultures on SMSB medium containing 250 μM DNT; some flasks also included 10 mM succinate, which had been observed to increase the growth rate of cells on plates. The results are summarized in Table 1. For strains DNT1 and PR7, the presence of succinate led to higher specific growth rates, higher cell densities, and thus higher overall DNT degradation rates (by factors of 1.4 and 1.9). However, for all three strains, the specific (per cell) rates of DNT consumption were found to be 6 to 22 times higher in the *absence* of succinate. These results provide strong motivation for the use of an immobilized cell

reactor: high densities of cells growing on succinate-free medium should result in a high volumetric productivity.

Strain PR7 was selected for use in the bioreactor because it had high overall and specific DNT consumption rates and because it accumulated only low levels of MNC for relatively short times.

maximum value of:

culture	μ (h ⁻¹)	overall DNT consumption rate (μ M/h)	specific DNT consumption rate (μ M/h-OD)
DNT1-S	0.04	9	1200
DNT1+S	0.09	13	200
PR7-S	0.04	13	1700
PR7+S	0.18	24	140
R34-S	0.08	11	2600
R34+S	0.09	11	120

Table 1. Results of screening experiment. Each flask was started with 5% (v/v) inoculum of a two-day culture on the same medium. "+S" and "-S" refer to the presence and absence of succinate in the medium, respectively.

B. DNT Biodegradation Kinetics

For a microorganism following the metabolic pathway shown in Figure 2, the rates of cell growth and DNT degradation are likely to be affected by the concentrations of DNT and nitrite, the presence of succinate (as a readily degradable carbon and energy source), and pH.

1. DNT Concentration

The effects of DNT level were evaluated in flask (batch) cultivations using suspended cells of strain PR7. The inoculum (10 vol%) came from a CSTR cultivation running at $D = 0.09 \text{ h}^{-1}$ and the flasks were incubated on a 250 RPM shaker at 32 °C.

The results of this experiment are shown in Figure 3 in terms of nitrate concentrations, which are stoichiometrically related to the concentration of DNT

degraded. Since cell concentrations were not measured in this experiment, no quantitative description of the kinetics can be made. However, two important qualitative observations are apparent: first, that exposure to higher initial levels of DNT lengthens the lag phase of the culture, and, second, that *Pseudomonas* PR7 is capable of degrading concentrations of DNT as high as 1500 μM .

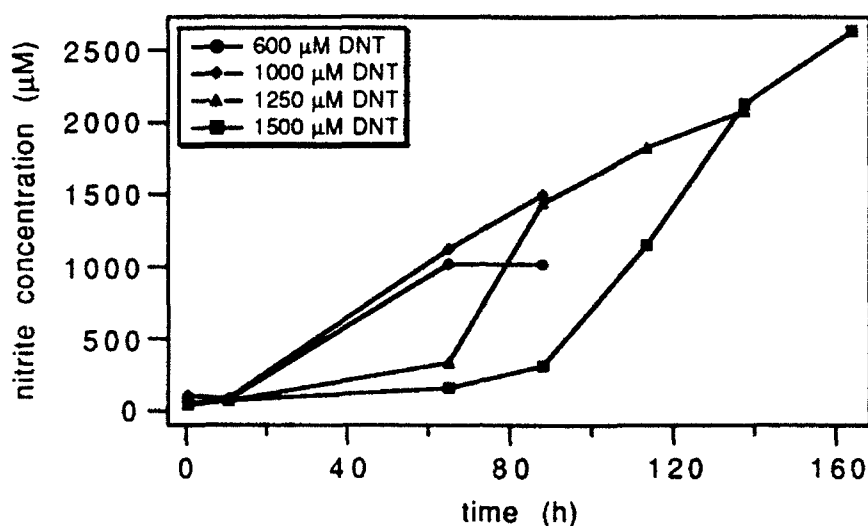


Figure 3. Nitrite released as a function of time during cultivations of *Pseudomonas* PR7 growing on SMSB + DNT (initial concentrations shown in legend).

2. Nitrite Concentration

Nitrite is a potent inhibitor of many microorganisms. Since the degradation of one mole of DNT releases two moles of nitrite, cultures degrading high levels of DNT could be subject to nitrite inhibition. Two sets of flask cultivations of *Pseudomonas* PR7 were performed in SMSB initially containing about 250 μM DNT and different levels of nitrite. After inoculation (5 vol%), the flasks were incubated at 32 $^{\circ}\text{C}$.

The results of the first set, in which the medium contained 10 mM succinate in addition to the DNT, are shown in Figure 4. It is apparent that nitrite inhibited growth (increased lag time, lower specific growth rate and cell yield) at 5000 μM ; however, the effects of 500 μM (initial) nitrite on growth are negligible in comparison to the control. Interestingly, high levels of nitrite did not lower the *specific* rates of DNT degradation.

The results of a similar nitrite inhibition study are shown in Figure 5. In this case, DNT was the only carbon source. High levels of nitrite had a stronger effect on

growth in this case than the previous, but again the specific rate of DNT degradation was unaffected by the addition of nitrite at these levels. For the degradation of DNT, the performance of a suspended cell reactor would suffer when high conversions were achieved, owing to the decrease in growth rate. However, an immobilized cell system, in which the performance can be independent of growth rate, would be unaffected.

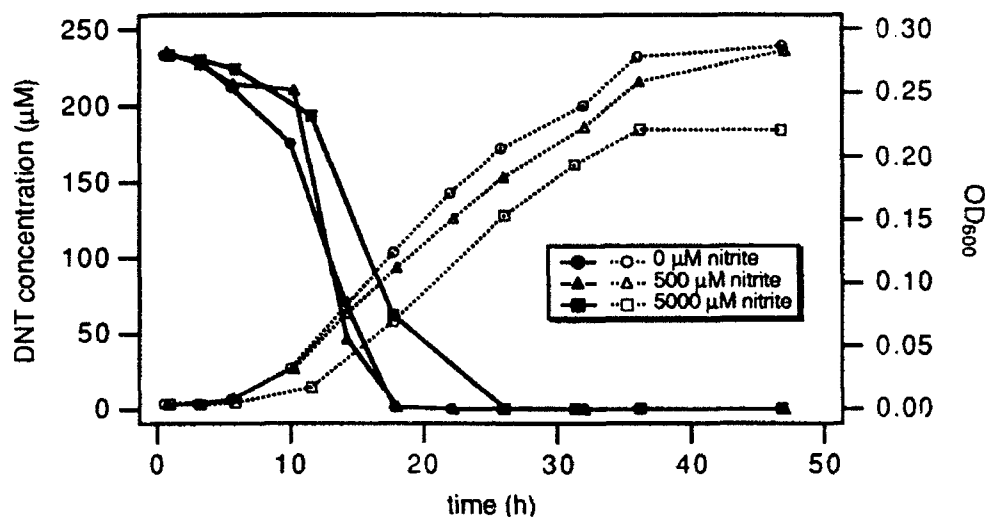


Figure 4. Results of batch nitrite inhibition studies with *Pseudomonas* PR7 (with succinate). Solid symbols: DNT concentration; open symbols: biomass (OD₆₀₀).

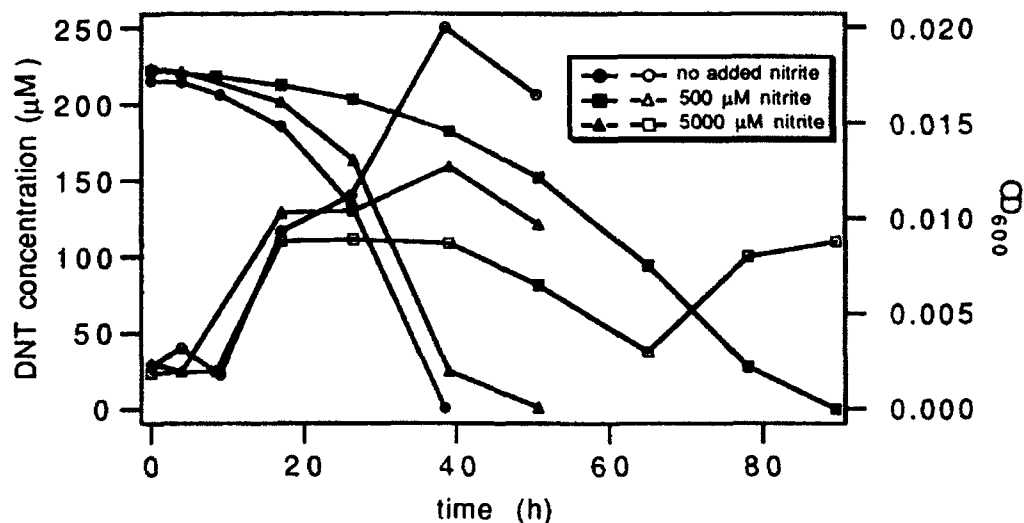


Figure 5. Results of batch nitrite inhibition studies with *Pseudomonas* PR7 (without succinate). Solid symbols: DNT concentration; open symbols: biomass (OD₆₀₀).

3. Succinate

Although the screening experiment had shown that the primary effect of succinate on the degradation of DNT was to increase the growth rate and thereby raise the volumetric degradation rate, it was not known if succinate and DNT were degraded sequentially or simultaneously. A flask culture (SMSB + 10 mM succinate + 250 μ M DNT, 32 °C, 5 vol% inoculum) was analyzed for both DNT and succinate levels. The results (Figure 6) indicate that both substrates are consumed at the same time, although succinate is degraded at a higher rate.

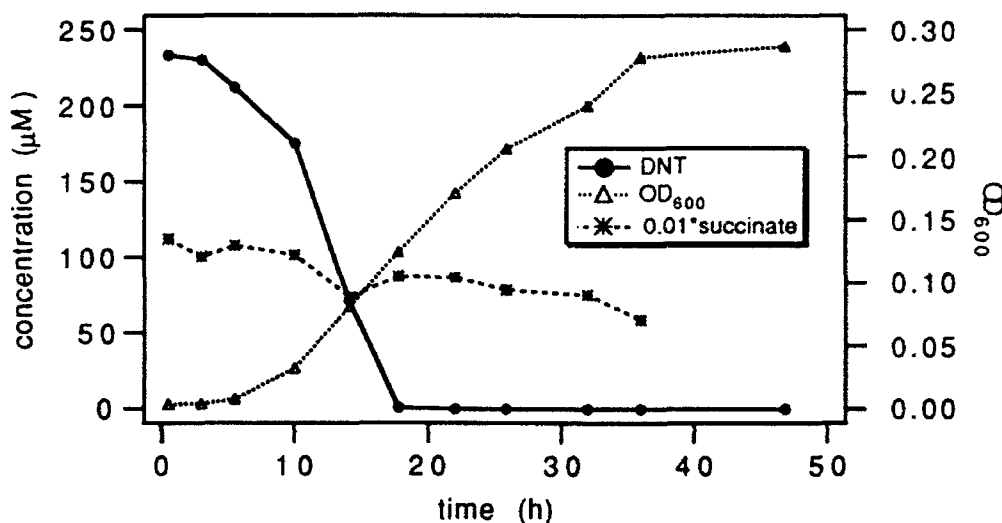


Figure 6. Succinate, DNT, and biomass concentrations during a batch cultivation of *Pseudomonas* PR7. Succinate values have been multiplied by 100 to fit the DNT scale.

4. pH

Since the nitrite groups are actually released as HNO_2 , lowering of the medium pH might occur during periods of high rates of DNT conversion. The effects of low pH were investigated by comparing two flask cultures, one buffered at pH 6.9 (the control) and the other at pH 5.75. Both flasks contained SMSB + 250 μ M DNT, were inoculated at 5 vol%, and were incubated at 32 °C. The results (Figure 7) show that cell growth is slower at pH 5.75. The specific DNT degradation rate, however, was unaffected.

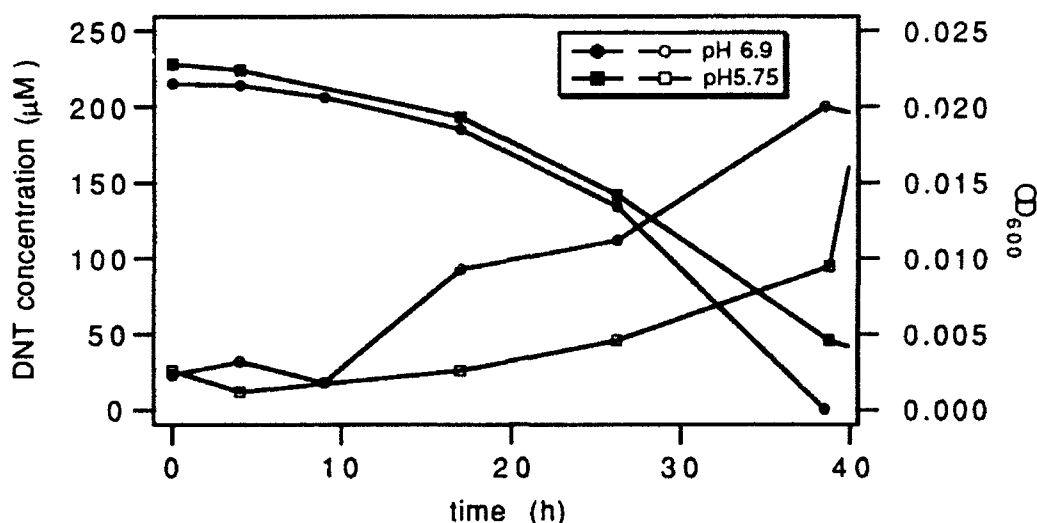


Figure 7. Effects of pH on batch DNT degradation by *Pseudomonas* PR7.

C. Immobilization Method

The immobilization method was surface attachment to Manville R635 diatomaceous earth pellets. These porous pellets are about 0.25 in. (diameter) by 0.5 in. (length); most pores are in the range of 8 – 40 μm [8]. This method was chosen for its simplicity and because strain PR7 had been observed to form biofilms.

The initial biomass loading in the bioreactor system was achieved by recirculating a growing PR7 culture through a column of pellets. Scanning electron micrographs were used to confirm the presence of cells on the surface of the pellets.

D. Bioreactor System

The bioreactor was initially configured as a continuous-flow, packed-bed, CSTR (Figure 8A). To assure that the system emulated a stirred reactor, the recirculation rate between the column and the mixing vessel was always at least 20 times greater than the flow rate through the complete system. Since no axial concentration gradients existed in the column, the activity of the cells was independent of their location in the column. Data interpretation and reactor scaleup were thus much simpler than in the case of a single-pass column, in which activity is a function of position in the column [2, 3].

This bioreactor system was constructed from a Pyrex chromatography column (Ace Glass), a Virtis fermenter vessel, and two peristaltic pumps. The tubing connecting

the two was primarily stainless steel, except for small pieces of silicon tubing in the pumps and at certain connections. The feed to the system contained DNT in SMSB.

After the recycle reactor system had been operated for 464 hours, a single pass (PFR) column was added to the system to achieve evaluate its ability to remove DNT from the effluent of the recycle (CSTR) system. This column was not inoculated directly but was allowed to capture cells in the CSTR effluent. A schematic diagram of the complete reactor system is shown in Figure 8B.

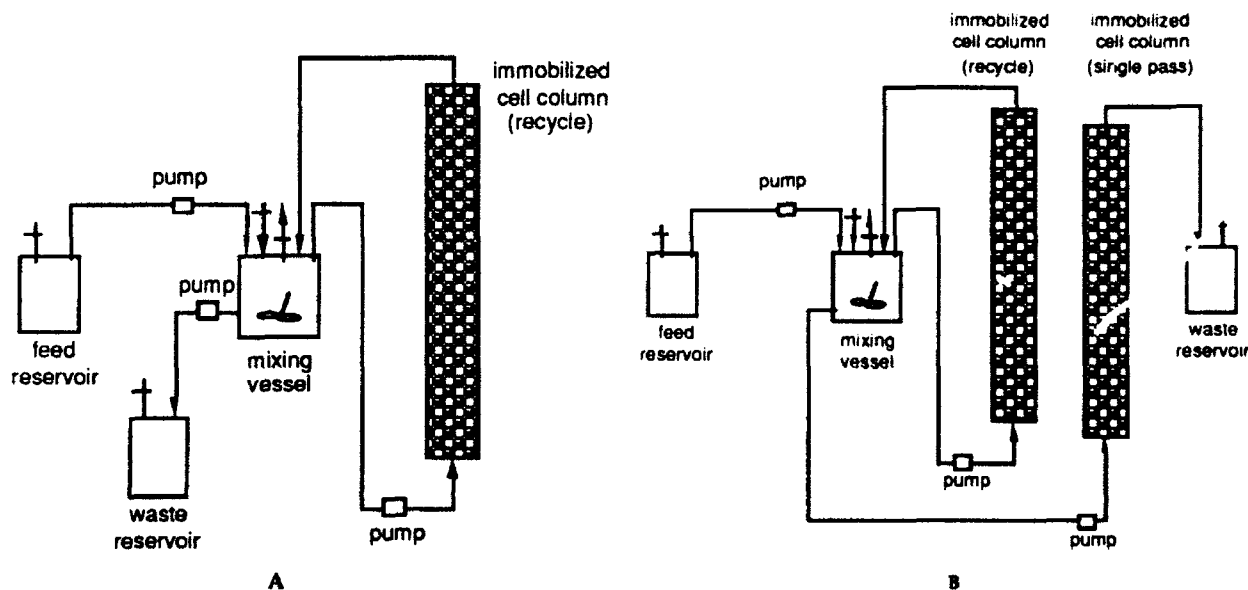


Figure 8. (A) Continuous-flow, packed-bed recycle reactor system that simulates CSTR operation. The total system liquid volume was 500 mL, the column dimensions were 60 cm (height) by 2.5 cm (diameter), the total (unpacked) column volume was 345 mL, and the liquid volume in the column was 135 mL. (B) Bioreactor system consisting of recycle (CSTR) and single-pass (PFR) immobilized cell columns (dimensions as above).

E. Characterization Of The Bioreactor System

Figure 9 shows results for the bioreactor system (CSTR portion) for the entire 712-h period of operation. During this run, variations in the feed composition, feed rate, and pH were imposed on the reactor to ascertain their effects (described below).

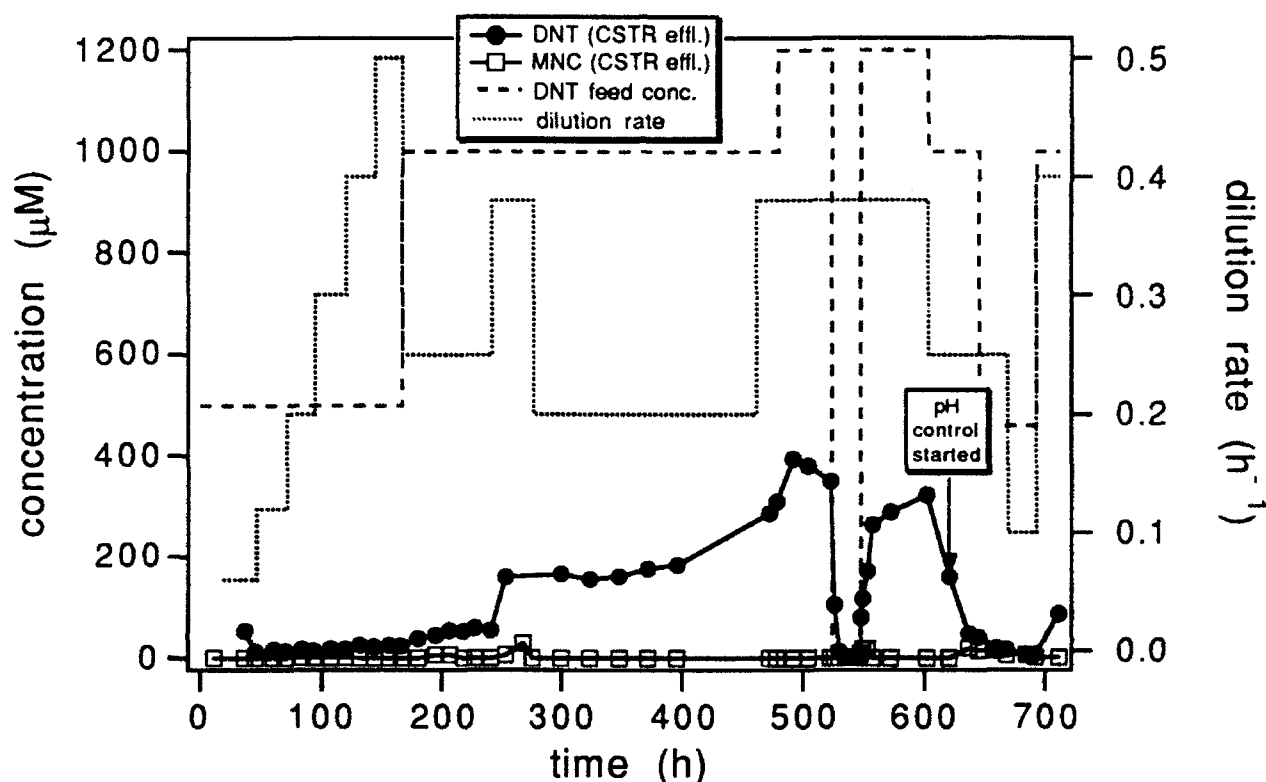


Figure 9. Results from the CSTR portion of the immobilized cell bioreactor.

1. Maximum Loading And Performance

In the first part of the bioreactor experiment, the loading ($\mu\text{mol DNT fed}/L_{\text{rxr}}/\text{h}$) was increased stepwise via increases in the dilution rate and feed concentration in order to test the performance of the system (CSTR portion) (Figure 10). As the loading increased during the first 250 h, the reactor performance increased proportionally and the concentration of DNT in the effluent remained relatively low. However, when the column loading was increased to $380 \mu\text{mol}/L_{\text{rxr}}/\text{h}$ at about 250 h, the performance did not increase to the extent expected and the effluent level of DNT increased. It appears

that some change occurred in the column at this time, since a return to a lower loading ($200 \mu\text{mol}/L_{\text{rxr}}/\text{h}$) did not produce the same level of performance. When the medium buffering was increased after 600 h to correct pH changes due to HNO_2 production (see below), performance rates returned to their original values.

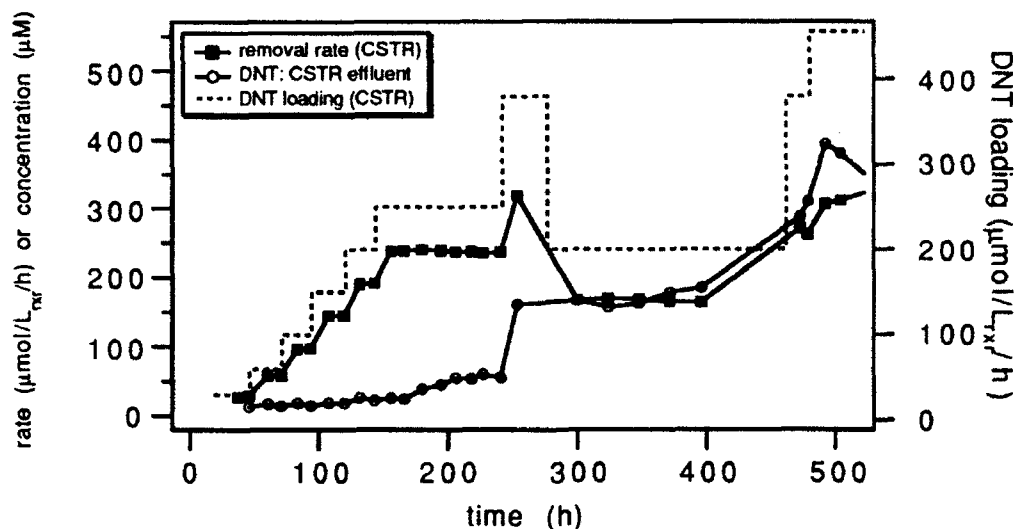


Figure 10. Performance (removal rate), loading, and effluent DNT concentration for the CSTR portion of the bioreactor system during the first 500 h of operation.

For scaleup purposes, these data can be rearranged as shown in Figure 11A. It is clear from this plot that the maximum performance of the CSTR portion of the bioreactor system is about $325 \mu\text{mol}/L_{\text{rxr}}/\text{h}$ for loadings greater than $325 \mu\text{mol}/L_{\text{rxr}}/\text{h}$. Although this removal rate value is based on the total CSTR system liquid volume (mixing vessel plus column), the value used for scaleup would be based on the total column volume since a very large column would require only a small mixing vessel. Thus, these removal rates should be multiplied by 1.45 to reflect this different volumetric basis, and the maximum performance based on the column volume is $470 \mu\text{mol}/L_{\text{rxr}}/\text{h}$. For comparison, an analogous plot for a suspended cell CSTR (also using PR7 and SMSB) is shown in Figure 11B. Although the removal rate vs. loading curve does not plateau in the loading range tested, it is clear that the maximum performance of this reactor is far less than that of the immobilized cell system.

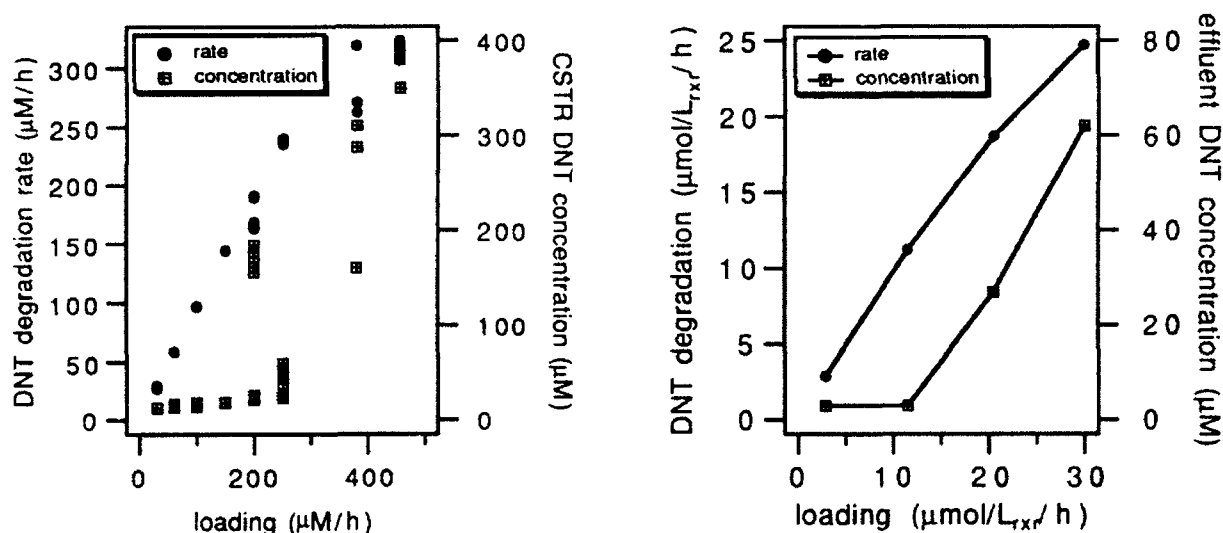


Figure 11. Performance and effluent DNT vs. loading curves for (A) the immobilized cell and (B) suspended cell CSTR systems.

The performance of the second (PFR) column was also assessed (Figure 12). Although this column did degrade DNT, it was not able to lower the DNT concentration in the CSTR effluent to a large extent. This was not due to oxygen limitation, since effluent from the second column contained dissolved oxygen at 97% of air saturation. Instead, poor cell loading is suspected to be the primary reason; this column was not inoculated as was the CSTR column but was connected to the process with the assumption that biomass shed from the CSTR system would colonize the diatomaceous support. Problems with low pH, due to high levels of DNT conversion, may have also contributed to the low performance of this column. This was corrected after 602 h by increasing the buffer capacity of the feed. Further study of the PFR portion of this bioreactor system is required.

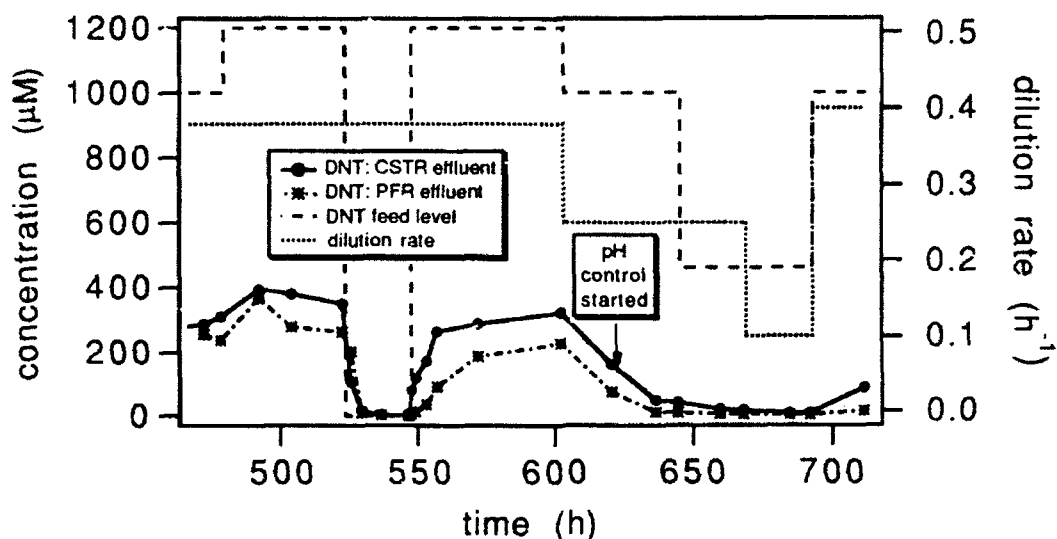


Figure 12. Data for the single-pass (PFR) column of the bioreactor system.

2. System Stability

An important characteristic of any reactor system is its ability to perform stably over time. The CSTR portion of the bioreactor was tested by maintaining a constant feed concentration and flow rate for more than 120 h. As Figure 13 indicates, this system was very stable under these conditions. The fact that the whole reactor system operated for more than 700 h under a variety of loadings also attests to its stability.

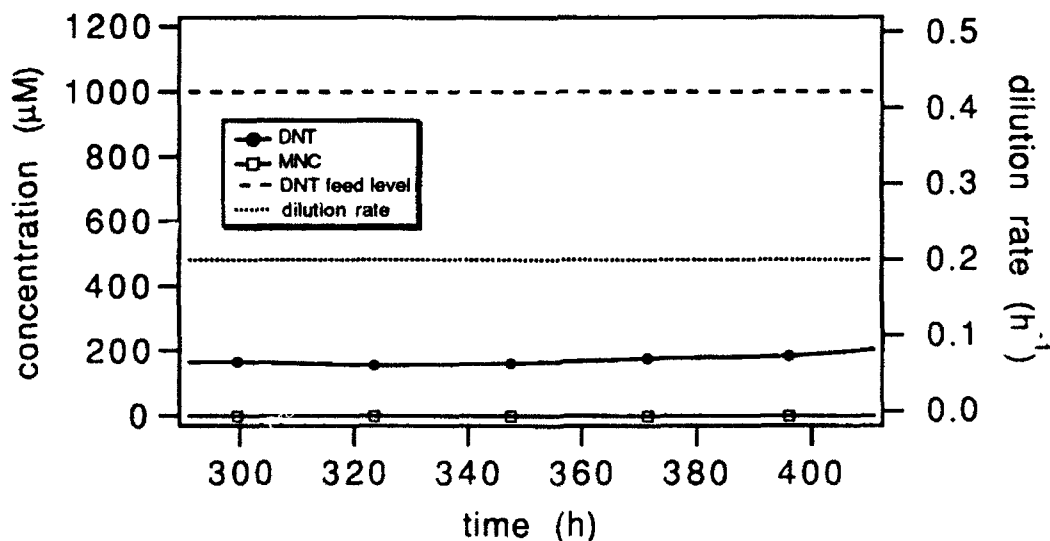


Figure 13. Results of stability test for the CSTR portion of the bioreactor system.

3. Effects Of Feed Fluctuations

Another important characteristic of a reactor system is its response to fluctuations in the loading. This is a special concern in this system, since DNT starvation could lead to uninduction of the necessary enzymes; when DNT feed levels increased, the reactor performance would be poor during the reinduction period. This bioreactor system was subjected to a perturbation in feed concentration: from 522.5 h until 547.5 h, the 1200 μM DNT feed was exchanged for one with no DNT. When the DNT feed was reintroduced, some MNC was observed in the effluent (indicating incomplete pathway induction) but no overshoot of DNT in the effluent occurred (Figure 14). Thus, the system withstood this severe loading fluctuation well.

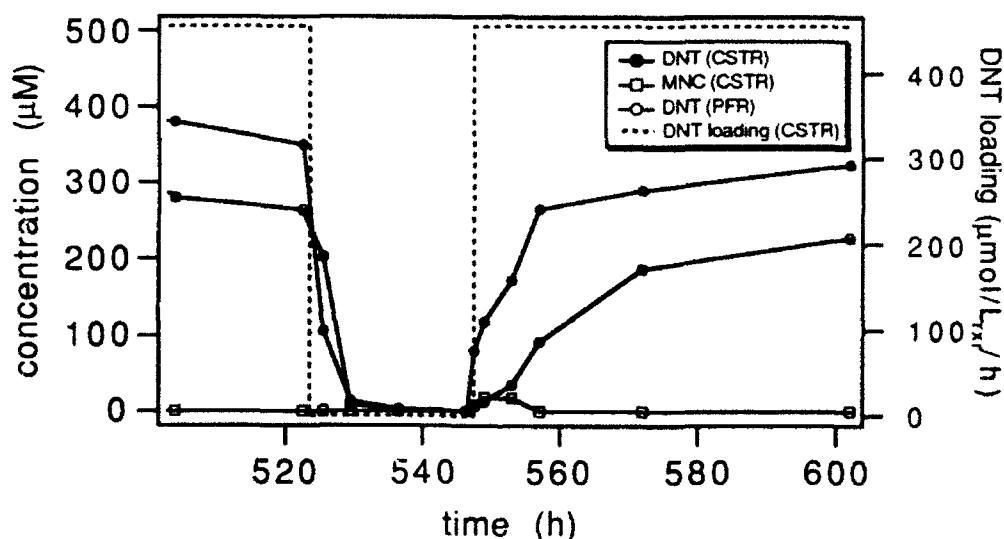


Figure 14. Results of feed fluctuation study. From 522.5 – 547.5 h, the feed concentration of DNT was lowered from 1200 μM to 0.

4. Response To Mixed Feed Streams

Since 2,4-DNT does not occur as a pure waste product, it was of interest to study the response of a column to a mixed waste stream containing DNT and 2, 4, 6-trinitrotoluene (TNT). PR7 will not grow on TNT alone but some preliminary experiments suggested that this strain would degrade TNT in the presence of DNT. During a 50-h period, the 2-column bioreactor system was fed a mixture consisting of 500 μM DNT and 100 μM TNT at two dilution rates. The results (Figure 15) show that both DNT and TNT were removed from the feed. The appearance of metabolites

(detected by HPLC) was further evidence that TNT was undergoing degradation. However, it was not clear from nitrite assays that TNT was degraded by a nitrite-releasing pathway similar to the one used by this organism for DNT; further study should be performed to determine this. An important result is that TNT did not affect the performance of the bioreactor toward DNT.

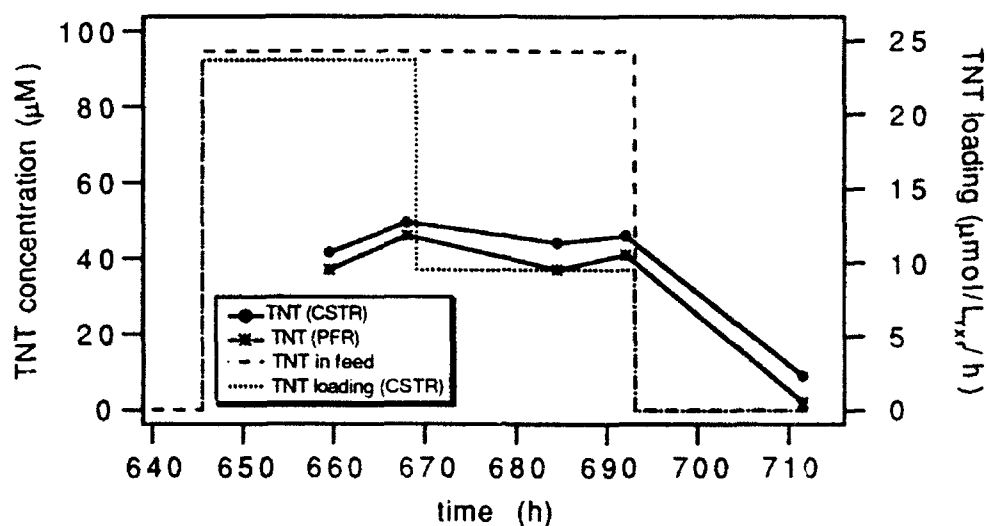


Figure 15. Results of the TNT/DNT mixed-feed study during the bioreactor experiment. Data for DNT consumption are not shown (see Figures 9 and 12).

SUMMARY

In the project, a bioreactor system for 2,4-dinitrotoluene degradation was developed and tested. This system was based on two packed columns of *Pseudomonas* PR7 attached to diatomaceous earth pellets; one column simulated a CSTR and the other was run in a single-pass mode. This bioreactor system operated continuously for more than 700 h. The CSTR portion of the system was studied in detail and was found to achieve degradation rates of 470 $\mu\text{mol}/\text{L}_{\text{rxr}}/\text{h}$ (based on total column volume), much higher than a comparable suspended cell system (25 $\mu\text{mol}/\text{L}/\text{h}$). Other tests showed that the system operation was stable and could withstand large fluctuations in DNT loading. Finally, the system was able to degrade DNT in the presence of a second contaminant (TNT) and also removed TNT, although at a slower rate.

REFERENCES

1. Leisinger, T. and W. Brunner, in: *Biotechnology*, vol. 8, H.-J. Rehm and G. Reed (eds.), VCH Publishers, Weinheim, 1986, pp. 475-513.
2. Bailey, J. E. and D. F. Ollis, *Biochemical Engineering Fundamentals*, 2nd Ed., McGraw-Hill, New York, 1986.
3. Mattiasson, B., in: *Immobilized Cells and Organelles*, vol. I, B. Mattiasson (ed.), CRC Press, Boca Raton, 1983, pp. 3-25.
4. Levenspiel, O., *The Chemical Reactor Omnibook*, Oregon State University, 1989.
5. Verschueren, K., *Handbook of Environmental Data on Organic Chemicals*, 2nd Ed., Van Nostrand Reinhold Co., New York, 1983, pp. 573-574.
6. Smibert, R. M. and N. R. Krieg, in: *Manual of Methods of General Bacteriology*, P. Gerhardt, R. G. E. Murray, R. N. Costilow, E.W. Nester, W. A. Woods, N. R. Krieg, and G. B. Phillips (eds.), American Society for Microbiology, Washington DC, 1981.
7. Spanggord, R. J., J. C. Spain, S. F. Nishino, K. E. Mortelmans, *Appl. Environ. Microbiol.* 57, 3200-3205 (1991).
8. Manville Corp. Data Sheet.

1992 USAF-RDL SUMMER FACULTY RESEARCH PROGRAM

Sponsored by the

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Conducted by

RESEARCH AND DEVELOPMENT LABORATORIES, INC.

FINAL REPORT

UNIAXIAL STRESS-STRAIN BEHAVIOR OF UNSATURATED SOILS
AT HIGH STRAIN RATES

Prepared by: George E. Veyera, Ph.D., Associate Professor

Department: Department of Civil and Environmental Engineering

University: University of Rhode Island

Research Location: HQ AFCESA/RACS
Tyndall AFB
Panama City, FL 32403-6001

USAF Researcher: Dr. C. Allen Ross

Date: 30 September 92

UNIAXIAL STRESS-STRAIN BEHAVIOR OF UNSATURATED SOILS AT HIGH STRAIN RATES

by

George E. Veyera, Ph.D.
Associate Professor
Department of Civil and Environmental Engineering
University of Rhode Island

ABSTRACT

The Split-Hopkinson Pressure Bar (SHPB) at AFCEA/RACS was used to study the uniaxial stress-strain behavior of compacted moist soils under one-dimensional, undrained, confined compression loading at high strain rates (1000/sec and 2000/sec). Three soils, Eglin sand, Tyndall sand and Ottawa 20-20 sand were tested. Each sand was compacted to a constant dry density at varying degrees of saturation between 0% (air dry) and 100% (fully saturated). General features of the uniaxial stress-strain results indicate: a) an initially steep loading portion associated with the initial rise in the loading pulse which appears to be strain rate independent; b) the slopes of each curve are about the same after the initial steep portion up to lock-up; c) the initial saturation affects the point at which lock-up occurs; and d) after lock-up the slope is approximately that of pure water. Lock-up refers to the sharp increase in the slope indicating a stiffening behavior at some compressive strain after the initial steep loading portion of the curve. The results suggest that the stress-strain response is dominated by the water phase from the lock-up strain and beyond, while the soil skeleton dominates the response from the start of loading up to the lock-up strain. It appears that there may be some strain rate effects, however, the data are insufficient to adequately demonstrate this and further investigations are required.

The research described in this report has demonstrated that the SHPB system is a viable technique for high strain rate dynamic geotechnical testing of unsaturated, saturated and dry soils, and provides a framework for conducting further studies using the SHPB with soils. While the saturation dependent uniaxial stress-strain behavior observed in this study has been theorized and hypothesized in the past by other researchers, these results appear to be the first detailed measurements of this phenomenon for undrained uniaxial confined compressive loadings at high strain rates. The results of studies such as that described herein will lead to a better fundamental understanding of the load transfer mechanisms in unsaturated soils and have direct applications to groundshock prediction techniques including stress transmission to structures. The research is important to the U.S. Air Force since there are currently no theoretical, empirical or numerical methods available for predicting the dynamic uniaxial stress-strain response of unsaturated soils from loading environments such as those produced by conventional weapons effects.

ACKNOWLEDGMENTS

I would like to thank the Air Force Systems Command, the Air Force Office of Scientific Research and the Air Force Civil Engineering Support Agency for sponsorship of this research work. Research and Development Laboratories provided assistance in all administrative and directional aspects of this program. I would like to especially thank Dr. C. Allen Ross for his friendship and his interest in the research topic. His continual support, encouragement and keen insight are very much appreciated. Special thanks are also due to the technical library at AFCESA which obtained numerous important reference items from a variety of sources. Thanks go to Mr. Clarence Schell who performed various tasks at the machine shop. In addition, I would also like to express my very sincere thanks to Mr. William S. Strickland at RACS for his hospitality and support while conducting this study. Finally, I would like to thank all of the many staff members at HQ AFCESA and at RACS for their camaraderie and friendship during the summer and for making my stay at Tyndall AFB so enjoyable.

UNIAXIAL STRESS-STRAIN RESPONSE OF UNSATURATED SOILS AT HIGH STRAIN RATES

by

George E. Veyera, Ph.D.

I. INTRODUCTION

Current empirical relationships for predicting soil pressure as a function of standoff distance from a buried explosive typically have a variation of $\pm 50\%$ or more and use material properties data based on conventional weapons effects in dry soils. However, most soils, whether naturally deposited or placed as select engineered fill, exist with moisture at saturations somewhere between 0% and 100%. The reaction of a structure to a specified loading can usually be determined, however, there are no theoretical, numerical or empirical methods available for predicting the groundshock energy arriving at a structure in unsaturated soils. This arises from the fact that dynamic load transfer mechanisms in soils are not well defined especially when moisture is present. In addition, there is little if any actual data available for the transient dynamic behavior of unsaturated soils, particularly at high strain rates.

The ability of a soil to transmit applied dynamic stresses (energy) is of particular interest to the U.S. Air Force with respect to military protective construction and survivability designs. Typical engineering analyses assume that little or no material property changes occur under dynamic loadings and in addition, analyses do not account for the effects of saturation (moisture conditions) on the stress-strain behavior of soils. This is primarily due to an incomplete understanding of soil behavior under transient loadings and uncertainties about field boundary conditions. Results from U.S. Air Force field and laboratory tests with explosive detonations in soils have shown that material property changes do in fact occur and that variations in soil stiffness (or compressibility) significantly affect both dynamic and static stress behavior. The research described in this report is important to the U.S. Air Force since there are currently no theoretical, empirical or numerical methods available for predicting the dynamic uniaxial stress-strain response of unsaturated soils from loading environments such as those produced by conventional weapons effects.

Recent research (2, 3, 9, 25-29) using the SHPB facility at AFCEA/RACS to study unsaturated soil behavior has shown that: (a) the presence and amount of moisture significantly affects the dynamic and static response of soil specimens; and (b) the amount of stress transmitted, stiffness, wave speed and compressibility in unsaturated soils varies with the amount of moisture present during compaction. Experimental evidence from a number of researchers suggests that such behavior for both dynamic and static loading conditions can be attributed to variations in soil compressibility and soil microstructure as a result of conditions during compaction including the compaction method used and the amount of moisture present during compaction (2, 3, 5, 9, 15, 16, 18-23, 25-29, 32). While the effects moisture on soil behavior as described above have been observed experimentally, a clear and concise explanation of the phenomenon is not currently available. This is

primarily due to the fact that the multiphase behavior of unsaturated soils, the interaction between the individual phases (air, water and solid), and the mechanics of load transfer mechanisms in soils are not well understood.

The research described herein was performed as a part of the 1992 Summer Faculty Research Program (SFRP) to investigate the undrained behavior of unsaturated soils subjected to dynamic confined uniaxial compression loading at high strain rates using the Split-Hopkinson Pressure Bar (SHPB) at AFCEA/RACS. The SHPB device can be used to examine the influence of selected parameters on the dynamic response of many engineering materials including soils. With reference to this research, the term "dynamic" defines large amplitude, high strain rate loadings as opposed to the low strain oscillatory frequency pulses used in wave attenuation studies. The results of studies such as that described herein will lead to a better fundamental understanding of the load transfer mechanisms and constitutive relationships for unsaturated soils and have direct applications to groundshock prediction techniques including stress transmission to structures.

II. RESEARCH OBJECTIVES

The primary objective of the summer research was to study the uniaxial stress-strain behavior of compacted moist soils under one-dimensional, undrained, confined compression loading at high strain rates using the SHPB at AFCEA/RACS. The SHPB testing system has been successfully used to evaluate metals, concrete, composites and foams at high rates of strain and the work described herein included the development of special equipment and techniques for using the SHPB with soils for which limited testing has been done. Particular emphasis was on examining the influence of saturation and strain rate on dynamic soil behavior. The research described is important to the U.S. Air Force since there are currently no theoretical, empirical or numerical methods available for predicting the dynamic stress-strain response of unsaturated soils to conventional weapons loading environments and such information is needed for more rational military protective construction and survivability designs.

III. BACKGROUND

Differences in the stress-strain response for dry and moist soils under both dynamic and static one-dimensional loading conditions have been observed experimentally by a number of researchers at various strain rates. Farr and Woods (6) accurately describe the current state of affairs with regards to the uniaxial stress-strain behavior of soils: "It has long been recognized that the one-dimensional or uniaxial strain response of most soils subjected to high intensity transient loads differs from the response measured under static conditions. As the time to peak pressure decreases, most soils exhibit a stiffening of the loading stress-strain response. The stiffening is usually referred to as a time or loading rate effect. Some researchers (14, 30) have suggested that, as the time to peak pressure approaches the submillisecond range, a drastic increase (up to tenfold) in the loading constrained modulus occurs for partially saturated granular soils under unconsolidated-undrained conditions. The existence of this effect has been the subject of much debate." A number of researchers have attempted to address the controversy surrounding the issue of strain rate effects in both dry and unsaturated soils.

Hendron (12) conducted one-dimensional confined compression tests on dry sands at very low loading rates, noting that the shape of the stress-strain curves became more S-shaped with increasing stress levels. Whitman et al. (31) also observed the typical S-shaped curve for dry sands and propose the following order of events with increasing applied load to describe this behavior: initially, deformations occur at the grain contact points as the individual grains deform; this is followed by a decrease in resistance to straining as the grains slide relative to each other; finally, an increase in resistance to straining occurs as the grains rearrange into a denser packing. Such behavior has been demonstrated theoretically from an analysis of a regular array of elastic spheres and has been shown to be dependent on the confinement boundary conditions (24). For a condition of zero lateral strain, the stress-strain curve exhibits a "strain hardening" effect with increasing stress (curve is concave towards the stress axis). In addition, Whitman et al. (31) noted an increase in modulus with decreasing time to peak loading for rise times ranging from the millisecond to seconds range attributed the stiffening response to strain rate effects.

High pressure static uniaxial confined compression tests were conducted on moist specimens of sandy silt at various dry densities by Hendron et al. (13). They observed increases in stiffness with increasing saturation (Figure 1) and from their results, concluded that the key variables in uniaxial stress-strain behavior are void ratio and saturation. Wu et al. (32) tested moist silty soils at small strains in the resonant column device and found a significant increase in the dynamic shearing modulus for specimens compacted moist at saturations in the range of from 5 to 20% (Figure 2). Whitman (30) indicated that rate effects become very important at submillisecond loading times and theorized that at high saturations, the pore phase of the soil is much stiffer than the soil skeleton (pore fluid compressibility dominates), while at lower saturations the skeleton matrix is stiffer than the pore phase (skeleton compressibility dominates). Jackson et al. (14) conducted unconsolidated-undrained one-dimensional confined compression tests on several different air dry soils at various loading rates and observed loading rate effects at submillisecond loading times, with the constrained modulus increasing by an order of magnitude in going from 0.1 to 1.0 ms rise time to peak stress. Farr and Woods (6) used a modified version of Jackson et al.'s (14) experimental apparatus to conduct similar tests on moist carbonate sand. They observed a progressive stiffening in the stress-strain response with faster loading times to peak stress and noted rate dependent effects occurred even at multi-second loading rates for the carbonate sand.

Limited research has also been conducted using the Split-Hopkinson Pressure Bar to study the uniaxial stress-strain behavior of compacted moist soil under one-dimensional confined compression loading at high strain rates (7, 8, 11). The results of these studies indicated that the uniaxial stress-strain response is primarily governed by the initial gas filled porosity of the soil specimen, and that strain-rate effects did not occur at strains less than the initial gas filled porosity. However, it should be noted that the specimen container boundary conditions were not for completely undrained conditions and therefore, this may have had significant effects on the results. The SHPB at AFESC/RACS (9, 25-29) has also been used to study stress transmission characteristics of unsaturated soils for conditions of undrained confined uniaxial compression. The

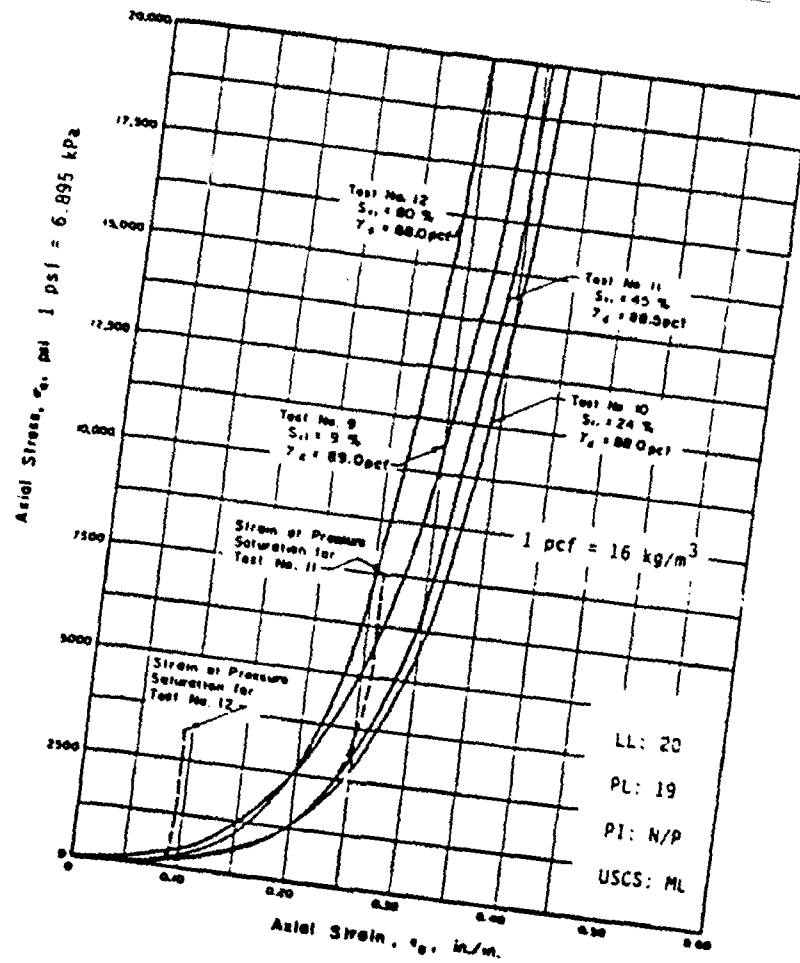


FIGURE 1. Uniaxial Compression Stress-Strain Response for Sandy Silt as a Function of Saturation (Hendron et al. (13)).

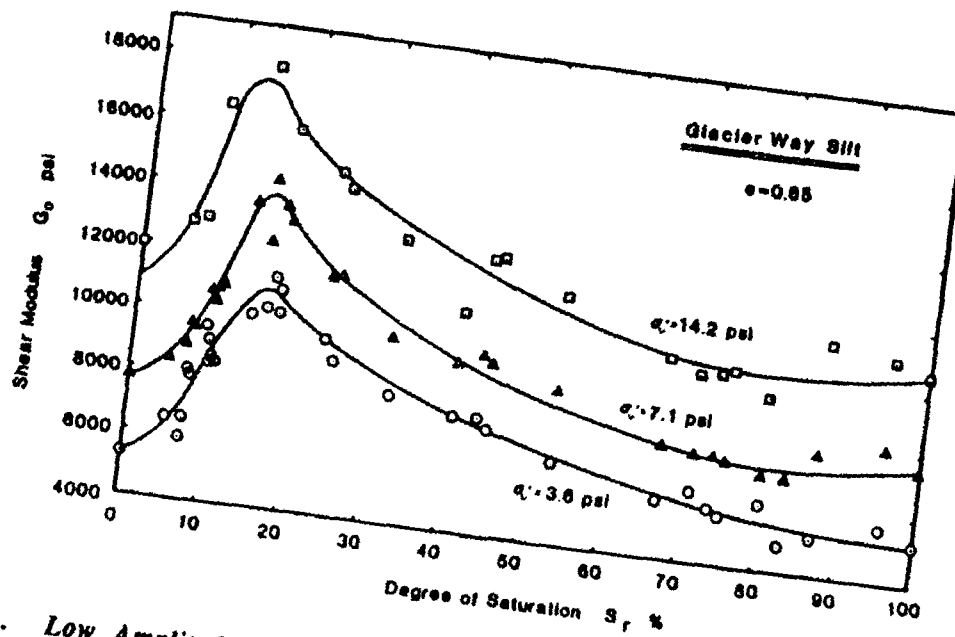


FIGURE 2. Low Amplitude Shear Modulus as a Function of Saturation for Glacier Way Silt (Wu et al., (32)).

quipment were used in the research described in this report to examine the undrained uniaxial compressive behavior of unsaturated soils at high strain rates. Detailed information about the SHPB device at AFCEA/RACS and results of recent investigations to study dynamic soil behavior can be found in the references cited above.

IV. EXPERIMENTAL INVESTIGATION:

A. Description of Soils Tested

Three different granular soils were tested using the AFCEA/RACS SHPB facility: Eglin sand (from Eglin AFB), Tyndall sand (from Tyndall AFB), and Ottawa 20-30 sand (commercially available from the Ottawa Silica Company). Representative soil samples were randomly obtained for analysis and testing from bulk quantities of each material using standard sample splitting procedures. The Eglin sand is a medium to fine, angular to subangular sand with about 7% fines; the Tyndall sand is a fine, uniform, subangular sand with no fines; and the Ottawa 20-30 sand is a uniformly graded, subrounded to rounded, medium sand with no fines. Various physical index properties data were obtained for each sand and the results are summarized in Table 1. A comparison of the grain size distributions is shown on Figure 3.

TABLE 1. Physical Properties of Eglin, Tyndall and Ottawa 20-30 Sands.

	Eglin	Tyndall	Ottawa 20-30
USCS Classification	SP-SM	SP	SP
Specific Gravity	2.65	2.65	2.65
D ₅₀ Particle Size (mm)	0.26	0.19	0.70 mm
^b C _u	3.41	1.18	1.40
^c C _c	1.29	0.95	1.03
^d Percent passing #100 sieve (%)	12	2	<1
^d Percent passing #200 sieve (%)	7	0	0
^e Maximum dry density (kg/m ³)	1,670	1,630	1,720
^f Minimum dry density (kg/m ³)	1,450	1,450	1,560
Maximum void ratio	0.817	0.817	0.705
Minimum void ratio	0.590	0.621	0.545
Note: ^aUnified Soil Classification System (1) ^bCoefficient of Uniformity ^cCoefficient of Curvature ^dU.S. Standard Sieve ^eASTM D4253 (1) ^fASTM D4254 (1)			

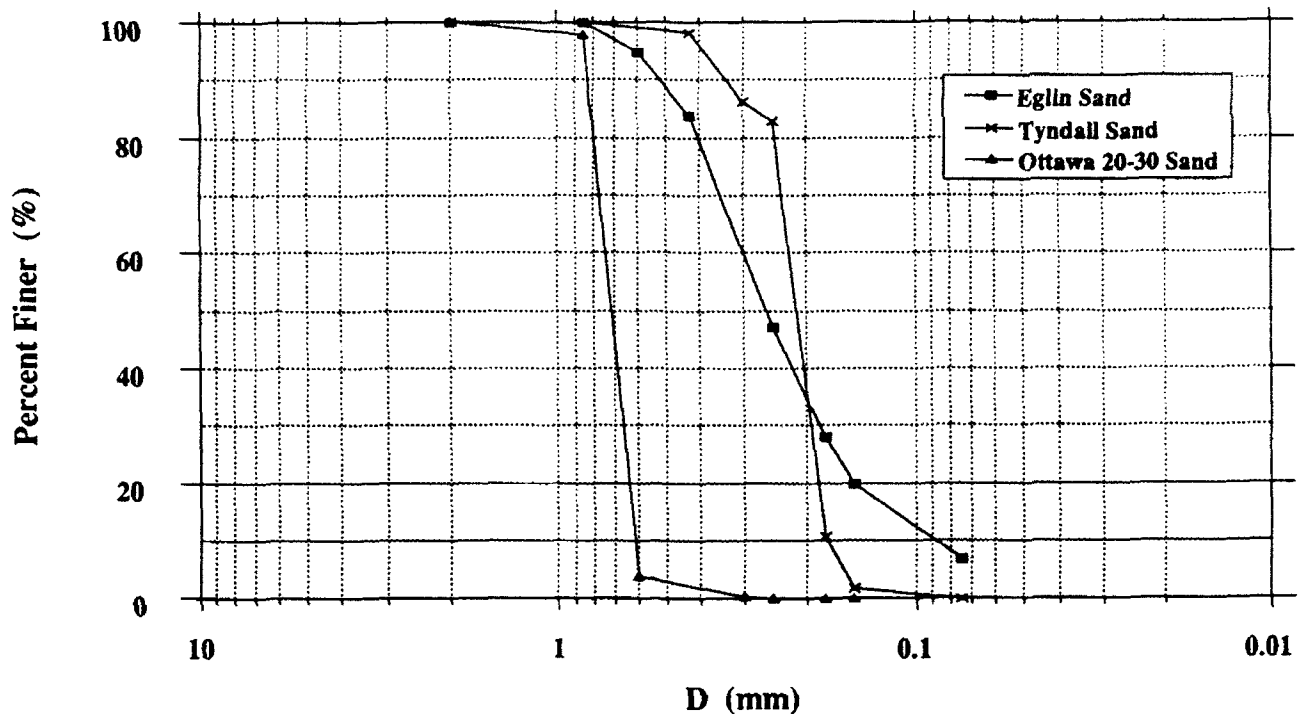


FIGURE 3. Grain Size Distributions for Eglin, Tyndall and Ottawa 20-30 Sands.

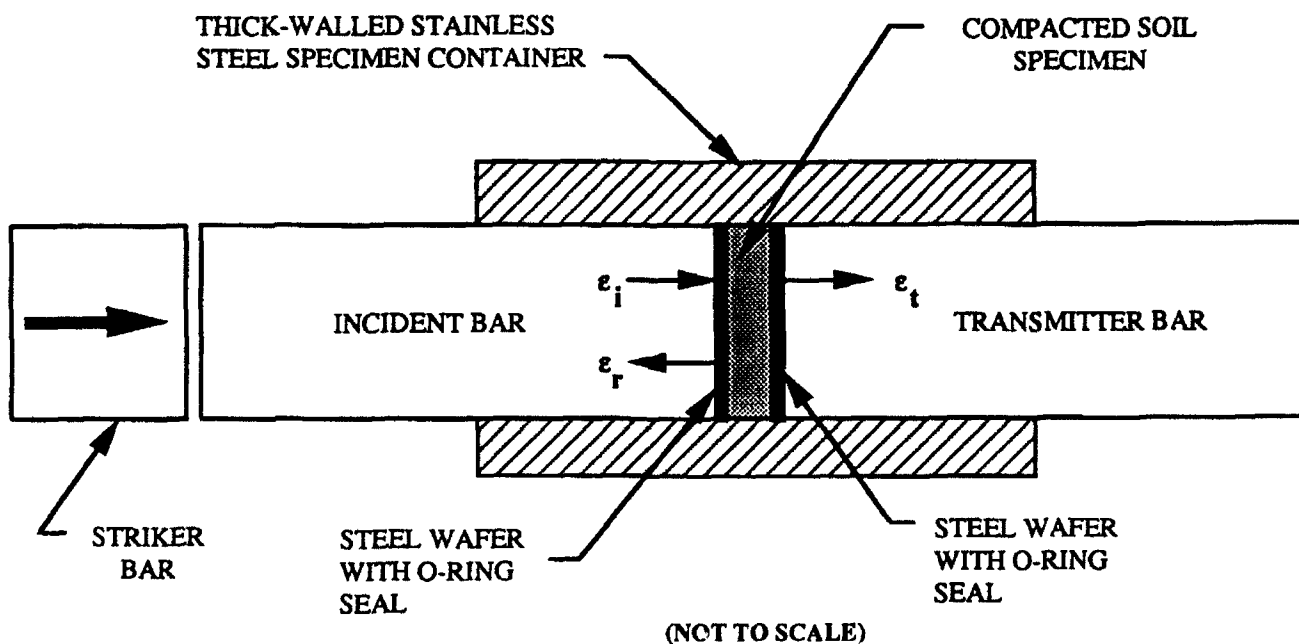


FIGURE 4. Schematic of Compacted Soil Specimen in the Split-Hopkinson Pressure Bar.

B. Preparation of Compacted Soil Specimens

Specimens of the Ottawa 20-30 sand were dynamically compacted to a constant dry density at varying degrees of saturation (different initial moisture contents) in a 7.62 cm (6.00 inches) long, 2.54 cm (1.00 inch) thick seamless stainless steel container which had an inside diameter of 5.08 cm (2.00 inches). The thick-walled stainless steel tube was used to simulate the one-dimensional, confined uniaxial loading condition typically encountered near explosive detonations in the field. Soil specimens are held in the container by two 0.635 cm (0.250 inch) thick stainless steel wafers fitted with o-ring seals used to prevent drainage of pore fluid during compaction and testing. One wafer is inserted prior to compaction and the other one after compaction is completed. Special care is taken when placing the second wafer after compaction to ensure full contact with the specimen. Figure 4 shows a compacted specimen prepared for testing in the SHPB.

A Standard Proctor hammer, ASTM D-698 (1), was used to consistently apply a controlled amount of compactive effort per impact to each soil specimen (7.5 Joules or 5.5 ft-lbs per impact). All test specimens were formed using a single individually compacted layer such that a final specimen length of 1.27 cm (0.50 inches) or 0.635 cm (0.25 inches) would be obtained at the maximum dry density for each soil. Saturations were varied from 0% to 100% for each soil.

In preparing moist specimens, the required amount of water for a given degree of saturation (at final compacted density) was added to the originally dry soil, thoroughly mixed in and then allowed to equilibrate before compacting. Although the dry density was constant for each specimen, the amount of compactive effort required varied with the amount of moisture (saturation). For specimens ranging from 0% to about 80% saturation, the tests were conducted on unsaturated specimens which implies that both continuous air and water phases exist in the soil (e.g., there are no isolated air or water pockets in the grain matrix). For most soils, this generally occurs at saturations less than about 85% (4). Fully saturated specimens (no air), were also prepared. For dry specimens, the dry soil was poured directly into the tube and then compacted.

C. Split-Hopkinson Pressure Bar

Figure 5 shows an overview of the AFCESA/RACS Split-Hopkinson Pressure Bar testing facility which consists of several separate but intimately related components: (a) a dynamic loading system which includes a nitrogen pressurized cannon used to fire 5.08 cm (2.00 inch) diameter stainless steel projectiles (striker) of varying lengths at the incident bar; (b) a stainless steel incident bar 5.08 cm (2.00 inch) in diameter and 3.66 m (12 feet) in length; (c) a stainless steel transmitter bar 5.08 cm (2.00 inch) in diameter and 3.35 m (11 feet) in length; (d) electronic strain gage instrumentation with power supplies and amplifiers; (e) a digital storage oscilloscope for data acquisition; and (f) a desktop computer for data reduction and analysis.

A series of tests was conducted on the Eglin, Tyndall and Ottawa 20-30 sands at varying degrees of saturation between 0% and 100% and at two different strain rates (Tyndall sand was only tested at one strain rate - 1000/sec). Strain rates were approximately 1000/sec and 2000/sec which were obtained by testing compacted specimens of two different lengths: 1.27 cm (0.50 inches) or 0.635 cm (0.25 inches), respectively. In all tests,

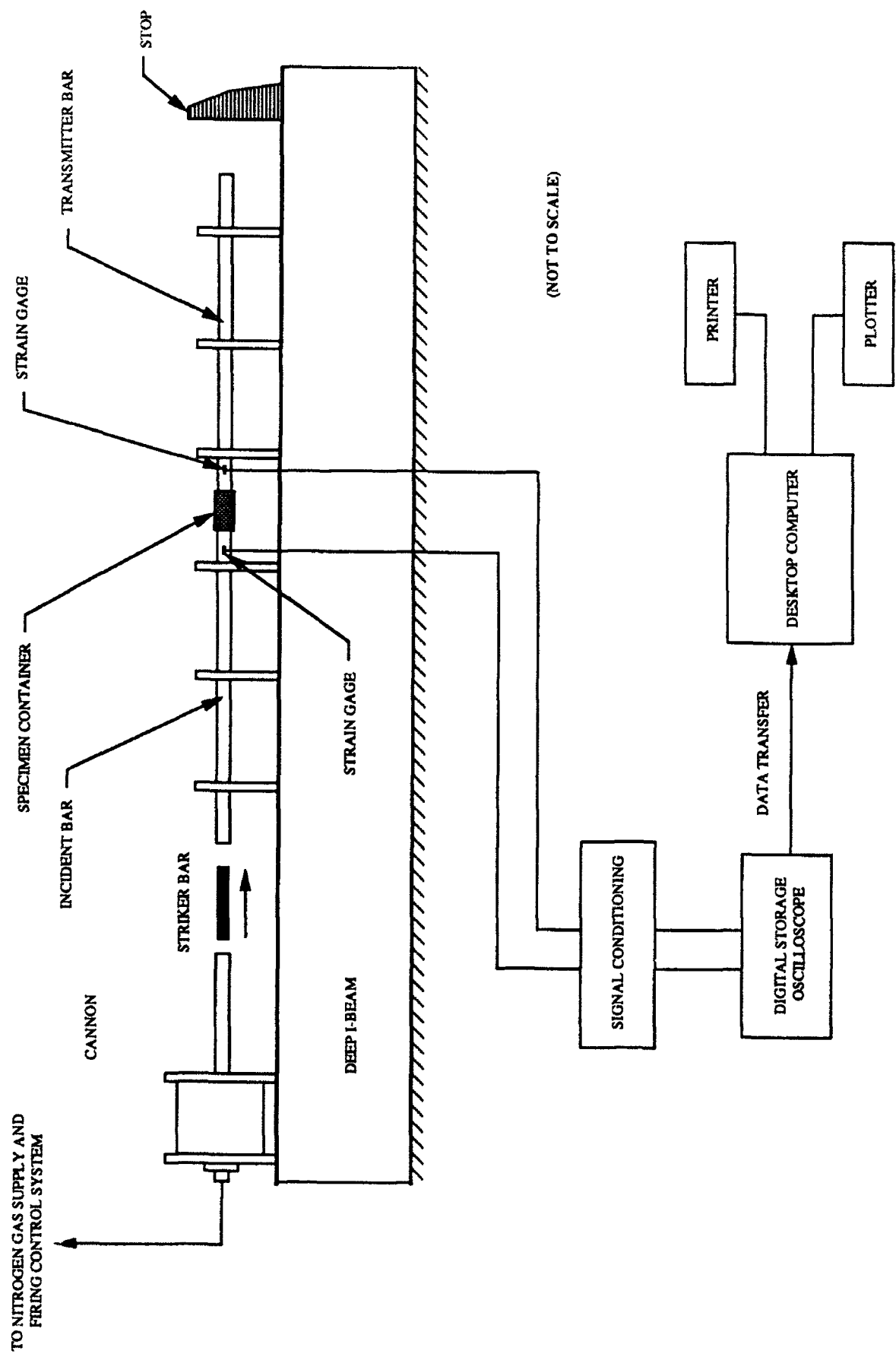


FIGURE 5. Schematic of Split-Hopkinson Pressure Bar Test Facility.

a 0.653 m (25.7 inch) long striker was fired at a cannon pressure of 690 kPa (100 psi), which produced a square wave input stress of approximately 225 MPa (37,000 psi) with a rise time to peak stress on the order of 50 microseconds and a 257 microsecond pulse width. The same cannon pressure, striker bar length and specimen container were used in all tests for each sand.

V. RESULTS AND DISCUSSION

The SHPB system provides measurements of the incident, reflected and transmitted strains in the incident and transmitter bars through electronic strain gage instrumentation and therefore, direct measurements of stress and strain within the specimen itself are not made (Figures 4 and 5). The average strain, average strain rate and average stress in the SHPB specimen can be determined from the strain gage data using the following relationships derived from elastic theory for one-dimensional stress wave propagation in a rod (17, 25):

$$\text{Average specimen strain:} \quad \epsilon_r = -\frac{2c_o}{L_o} \int_0^t \epsilon_i dt \quad (\text{Eq. 1})$$

$$\text{Average specimen strain rate:} \quad \dot{\epsilon}_r = -\frac{2c_o}{L_o} \epsilon_i \quad (\text{Eq. 2})$$

$$\text{Average specimen stress:} \quad \sigma_r = E \epsilon_i \quad (\text{Eq. 3})$$

where: L_o is the initial specimen length, c_o is the wave propagation velocity of the incident and transmitter bars, E is Young's modulus, and ϵ_i , ϵ_r and ϵ_t are the incident, reflected and transmitted strains, respectively.

The derivation of Eqns. 1, 2 and 3, assumes: a) the incident and transmitter bars are of same material (i.e., they have the same wave speed); b) loading stresses are in the elastic range of the bars; c) a uniform one-dimensional stress state is developed in the specimen; d) the forces on each end of the specimen are equal; and e) the cross-sectional areas of the bars and the specimen are equal. These equations are used in analyzing the raw data to develop the uniaxial stress-strain curves (Figures 6 and 7). The data analysis includes a dispersion correction to account for wave spreading in the bars and an FFT using 17 point smoothing ($n=17$).

Figures 8, 9 and 10 show the uniaxial stress-strain results for the Eglin, Tyndall and Ottawa 20-30 sands, respectively. Data were obtained for the Eglin and Ottawa 20-30 sands at strain rates of 1000/sec and 2000/sec, while the Tyndall sand data were only obtained at a strain rate of 1000/sec. In addition, there were some difficulties in obtaining reliable data for the Ottawa 20-30 sand at 100% saturation, therefore those data have not been included. There are some general features observable in the stress-strain data: a) an initially steep loading portion of the curves which is probably associated with the initial rise in the loading pulse and appears to be strain rate independent; b) the slopes of each curve are about the same after the initial steep portion up to the lock-up strain; c) the initial saturation affects the point at which lock-up occurs (the higher the initial saturation,

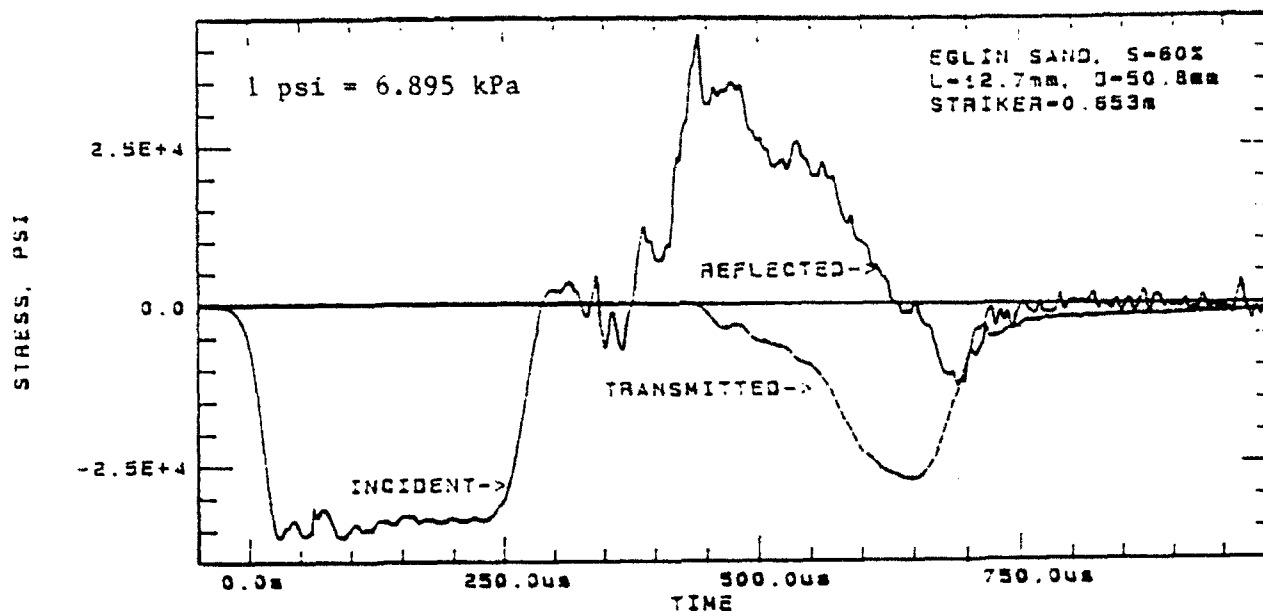


FIGURE 6. Typical SHPB Data for Incident, Reflected and Transmitted Stresses as a Function of Time for Uniaxial Compressive Loading.

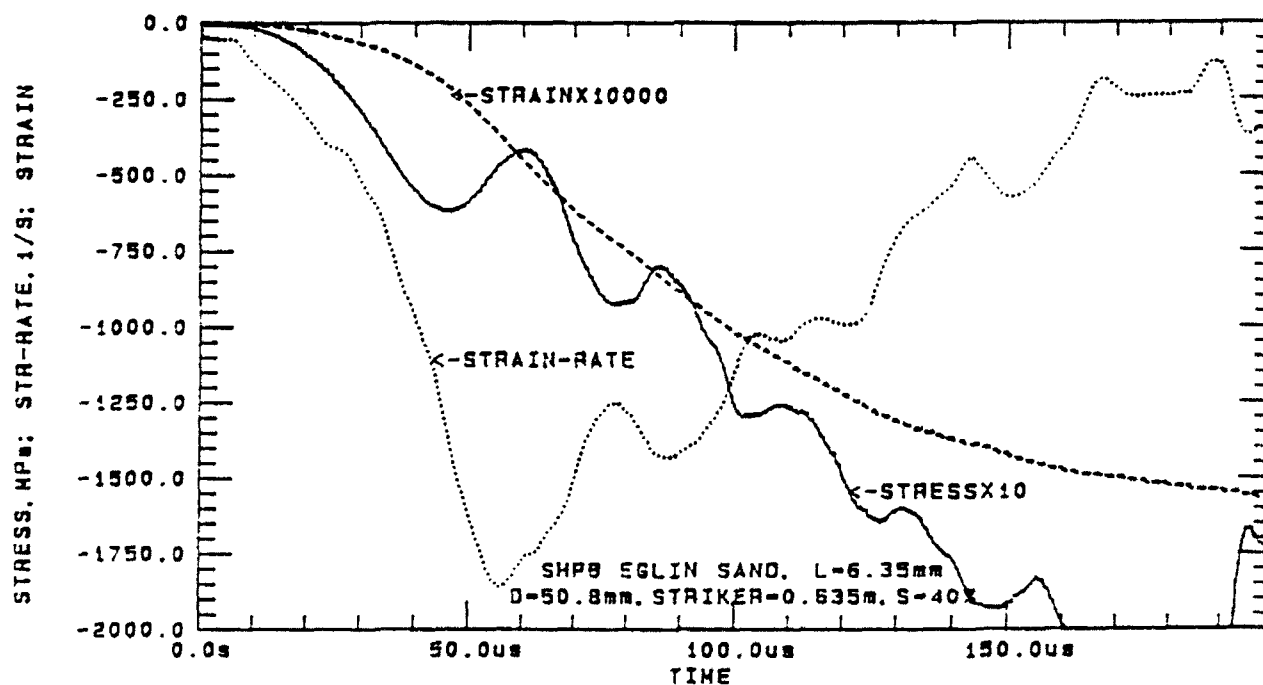


FIGURE 7. Typical SHPB Data for Strain, Strain Rate and Stress as a Function of Time for Uniaxial Compressive Loading.

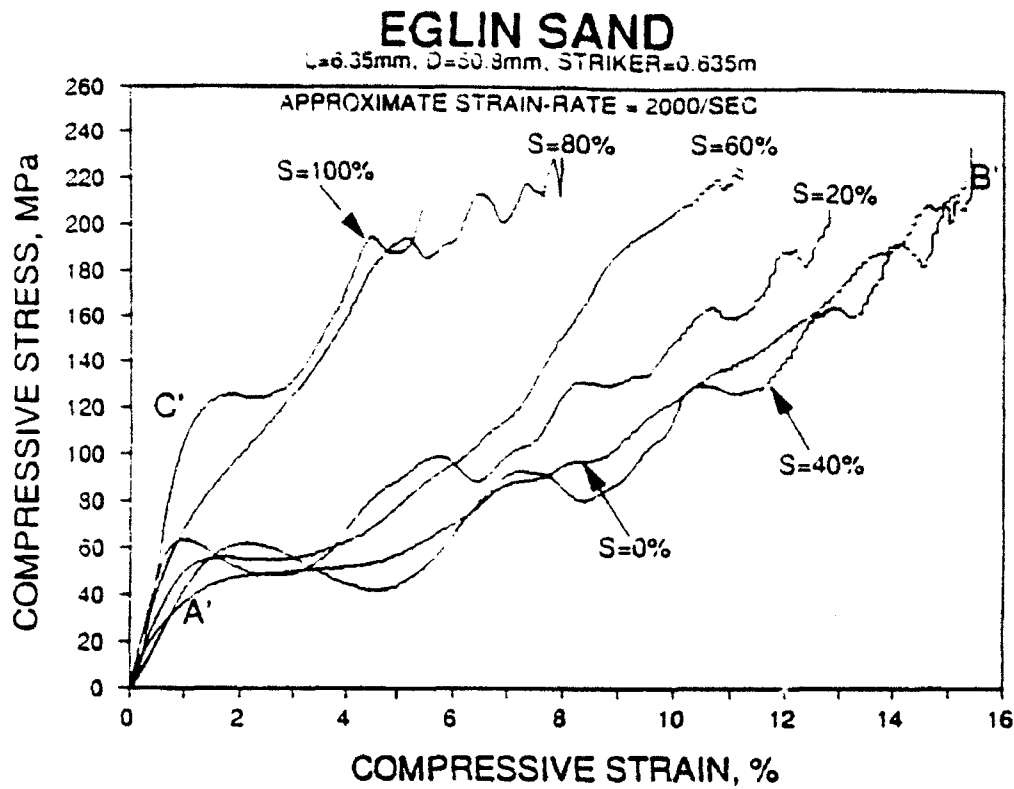


FIGURE 8a. SHPB Uniaxial Undrained Compressive Stress-Strain Response for Eglin Sand as a Function of Saturation (Strain Rate is Approximately 1000/sec).

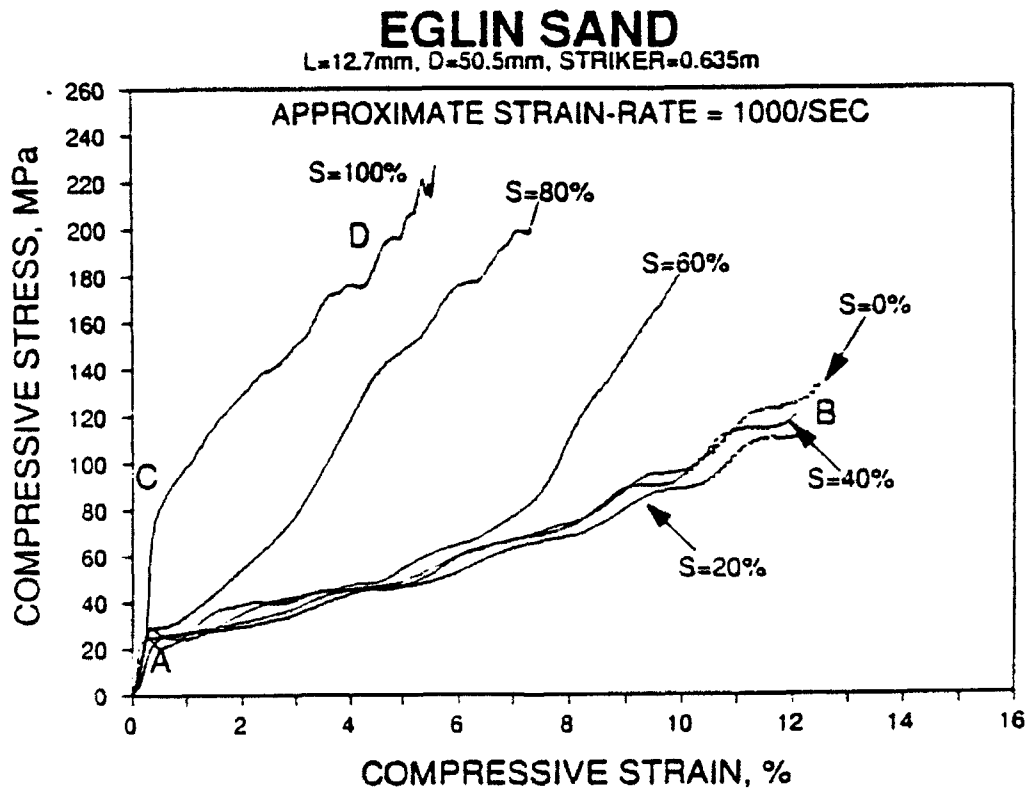


FIGURE 8b. SHPB Uniaxial Undrained Compressive Stress-Strain Response for Eglin Sand as a Function of Saturation (Strain Rate is Approximately 2000/sec).

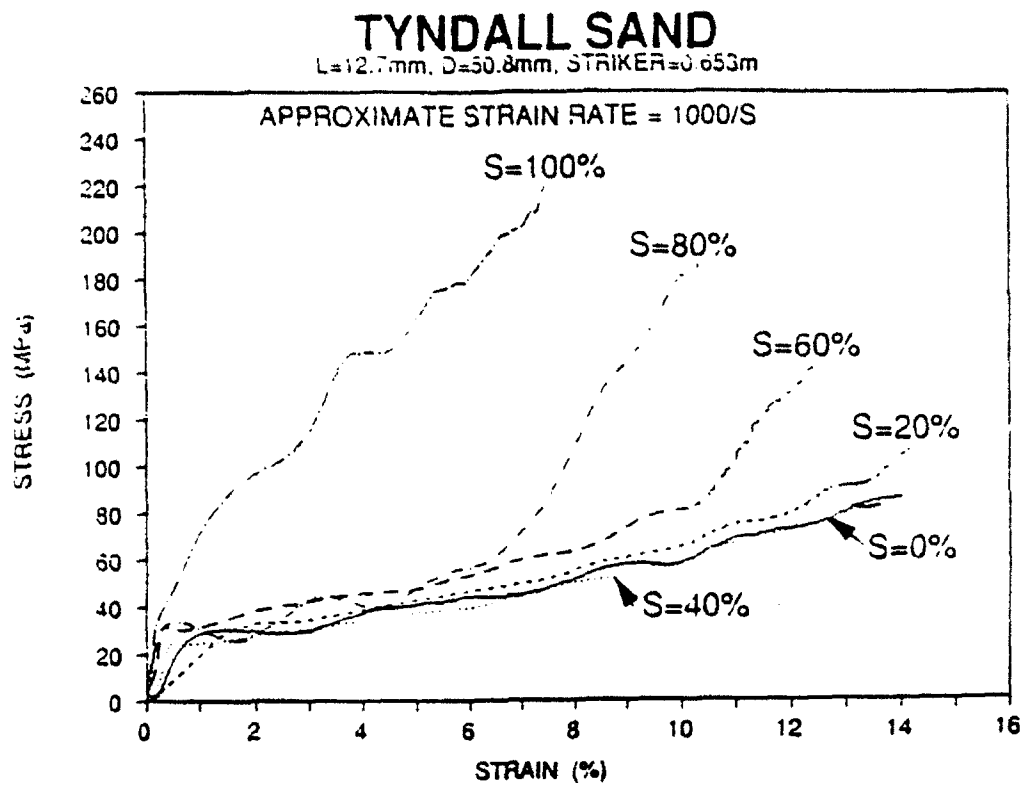


FIGURE 9. SHPB Uniaxial Undrained Compressive Stress-Strain Response for Tyndall Sand as a Function of Saturation (Strain Rate is Approximately 1000/sec).

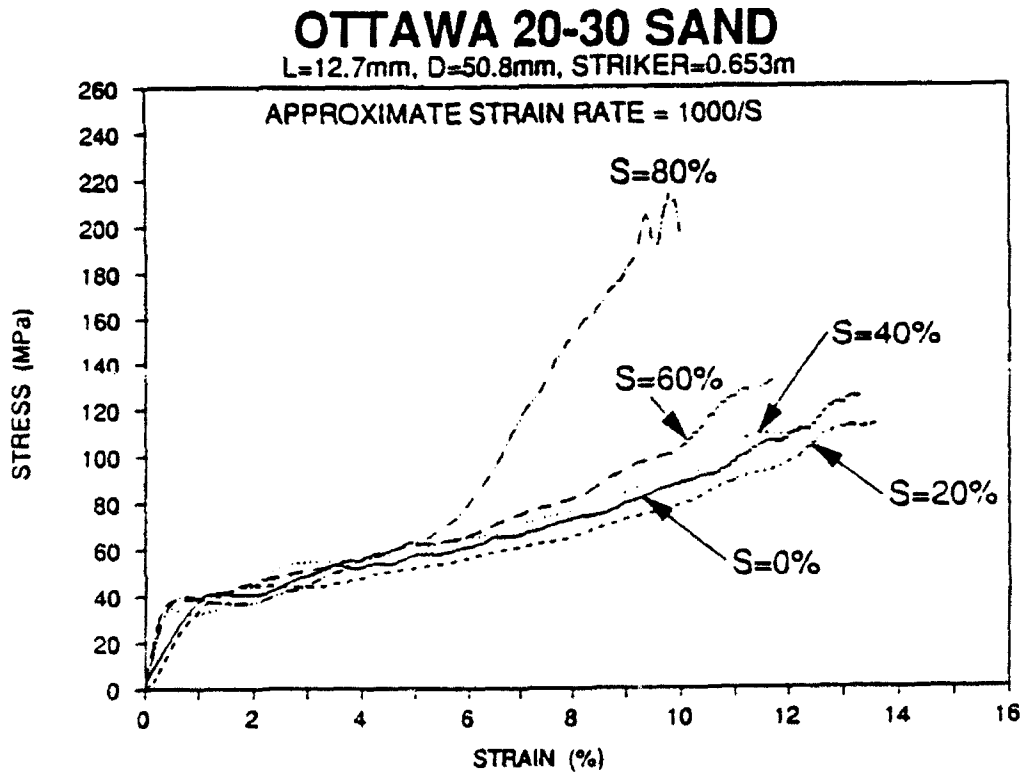


FIGURE 10a. SHPB Uniaxial Undrained Compressive Stress-Strain Response for Ottawa 20-30 Sand as a Function of Saturation (Strain Rate is Approximately 1000/sec).

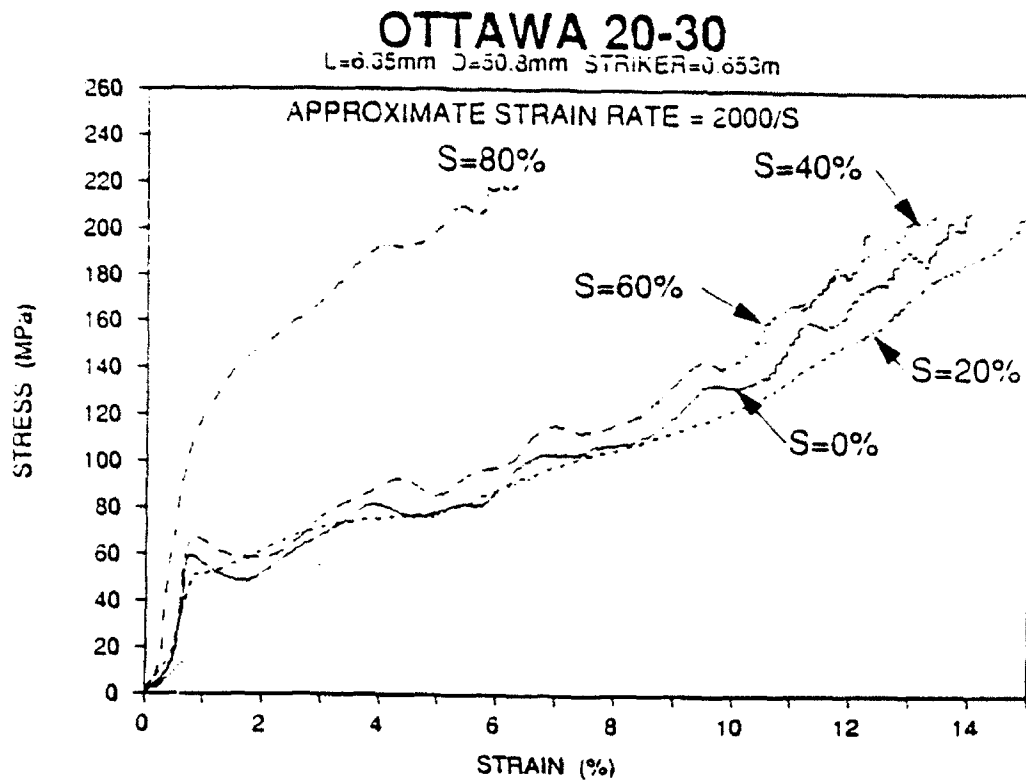


FIGURE 10b. SHPB Uniaxial Undrained Compressive Stress-Strain Response for Ottawa 20-30 Sand as a Function of Saturation (Strain Rate is Approximately 2000/sec).

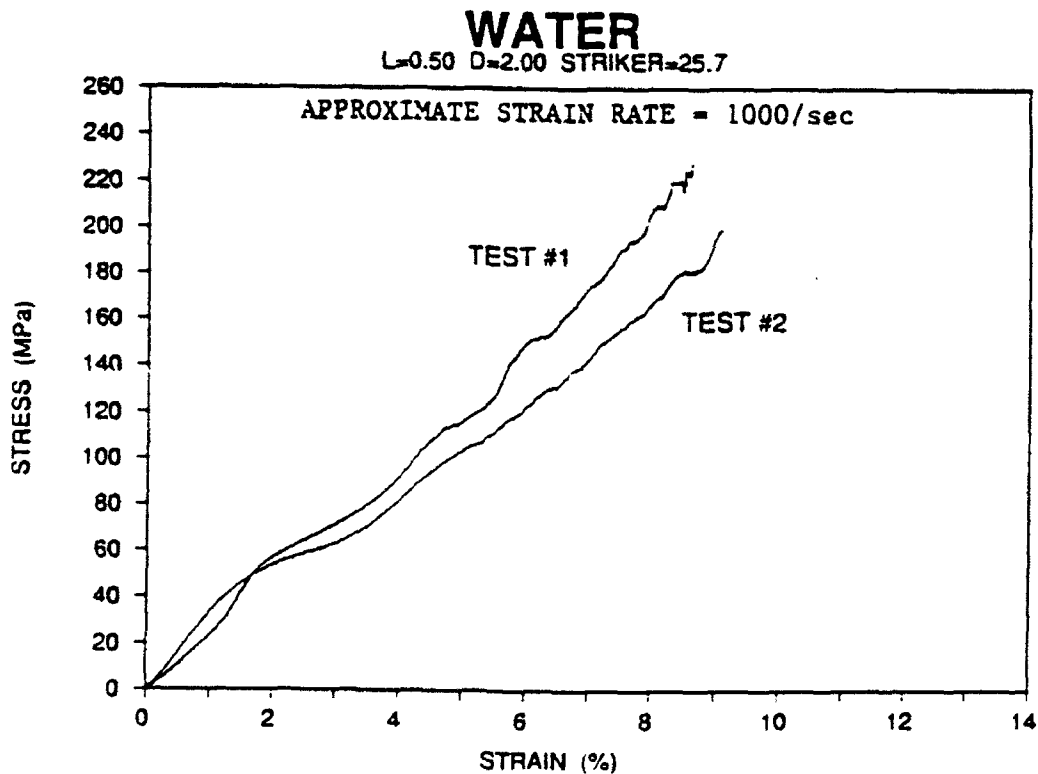


FIGURE 11. SHPB Uniaxial Undrained Compressive Stress-Strain Response for Water (Strain Rate is Approximately 1000/sec).

the smaller the strain required); and d) after lock-up the curves exhibit a stiffening response with a slope approximately that of pure water (Figure 11). In comparing the test results at different strain rates, it appears that there may be some strain rate effects, however, the data are insufficient to adequately demonstrate this and further investigations are required. The data at a strain rate of 2000/sec show increasing dispersion (waviness) in the curves for the Eglin sand, being much less for the Ottawa 20-30 sand. The dispersion effect is most likely due to the significant differences in particle characteristics of each soil (grain size, shape and size distribution), and it may also be that the small specimen length is approaching some limiting value in terms of particle size relative to loading pulse length.

In comparing the three different soils together it can be seen that: a) the slopes of the curves after the initial steep portion are nearly identical regardless of the initial saturation and soil type; b) the initiation of the lock-up strain is somewhat different for each soil at the same saturation; c) lock-up was not developed in the Ottawa 20-30 sand below 80% saturation. Differences in the basic features between the curves when comparing the three different soils are most likely due to differences in grain size characteristics (i.e., size and distribution).

The term "lock-up" as used herein refers to the sharp increase in the slope indicating a stiffening behavior of the stress-strain curve at some compressive strain (after the initial loading portion of the curve). Since the slope is approximately that obtained for SHPB tests conducted on pure water (the saturated soil-water mixture slopes are slightly greater than that for water due to density differences), the lock-up strain represents a condition of full saturation in the soil due a reduction in void space. Therefore, the results suggest that the stress-strain response is dominated by the water phase (saturated soil-water mixture) from the initiation of the lock-up strain and beyond, while the soil skeleton (unsaturated soil-water mixture or soil-air mixture if dry) dominates the response from the start of loading up to the lock-up strain. Lock-up was not developed at lower saturations even at large strains, which indicates that insufficient pore space reduction occurred (i.e., not enough compressive strain was developed). Therefore, larger amplitude stress loadings (higher compressive strains) with longer pulse lengths need to be applied so that the stress-strain response at the lower saturations can be determined.

Table 2 shows a comparison of the approximate measured lock-up strains estimated from the SHPB data (Figures 8, 9 and 10) and the theoretical compressive strain required to reach full saturation (i.e., lock-up). The theoretical calculations only account for the amount of air-filled void space in each soil based on initial void space and initial saturation. The results generally do not agree with the measured compressive strains at lock-up. However, the interaction of the various components of the soil-water mixtures are not accounted for in these calculations. Since the soil-water mixture represents a multi-phase material, a complete description of the problem becomes very complex since in general there are four different interacting components contributing to the bulk response: a) the soil skeleton; b) the pore air; c) the pore water; and d) the individual grain stiffnesses. While the individual stress-strain response of each component can be determined separately, it is their interrelationship that determines the overall behavior. A further complicating factor is that the dominance of any one (or combination) of these components changes depending upon the initial conditions and those during

loading (i.e., strain magnitude). However, these interrelationships are not well defined or understood, particularly for transient dynamic loading conditions.

Even though this investigation is not exhaustive in extent, it does shed some new light (and poses new questions) on the undrained dynamic behavior of unsaturated soils at high strain rates and represents an important first step towards establishing an understanding of this phenomenon. In addition, this research has also demonstrated that the SHPB system is a viable technique for high strain rate dynamic geotechnical testing of unsaturated, saturated and dry soils, and provides a framework for conducting further studies using the SHPB with soils. While the saturation dependent uniaxial stress-strain behavior observed in this study has been theorized and hypothesized in the past by other researchers, these results appear to be the first detailed measurements of this phenomenon for undrained uniaxial confined compressive loadings at high strain rates.

TABLE 2. Comparison of Approximate Measured Compressive Strains at Lock-Up and Theoretical Compressive Strains to Reach S=100% Based on Initial Void Ratio and Initial Saturation.

S (%)	EGLIN ^a			TYNDALL ^b			OTTAWA 20-30 ^c		
	e ^d	e ^e	e ^f	e ^d	e ^e	e ^f	e ^d	e ^e	e ^f
0	33.8	(g)	(g)	39.5	(g)	(i)	35.3	(g)	(g)
20	27.0	(g)	(g)	31.6	(g)	(i)	28.2	(g)	(g)
40	20.3	(g)	(g)	26.7	(g)	(i)	21.2	(g)	(g)
60	13.5	7.5	6.0	15.8	10.5	(i)	14.1	(g)	(g)
80	6.8	3.0	2.0	7.9	6.5	(i)	7.0	6.0	1.5
100	0.0	0.5	1.0	0.0	0.5	(i)	0.0	? (h)	? (h)

- Note:
- a Eglin sand at initial void ratio = 0.510
 - b Tyndall sand at initial void ratio = 0.654
 - c Ottawa 20-30 sand at initial void ratio = 0.545
 - d Theoretical compressive strain required to obtain S=100% based on initial void space and saturation. For S=0%, this represents a condition of zero air voids
 - e Approximate lock-up compressive strain from SHPB data for strain rate = 1000/sec.
 - f Approximate lock-up compressive strain from SHPB data for strain rate = 2000/sec.
 - g Unable to obtain lock-up strains using the 0.653 m (25.7 inch) projectile at 690 kPa (100 psi).
 - h Unable to obtain reliable data for Ottawa 20-30 sand at S=100%.
 - i Data not obtained for Tyndall sand at this strain rate.

VI. RECOMMENDATIONS

The understanding of load transfer mechanisms in unsaturated soils is very limited at present. Further investigations are necessary and will provide important information to the U.S. Air Force with respect to military protective construction and survivability designs. Recommendations for further research include:

1) The fundamental aspects of load transfer mechanisms, material phase interactions and the effects of boundary conditions in compacted unsaturated soils should be examined for both static and dynamic loading conditions in a comprehensive experimental testing program. Such a program should include the following:

- a) A more detailed investigation should be undertaken to study strain rate effects in unsaturated soils using the SHPB. In addition, centrifuge tests using scaled explosive charges should also be conducted. Based on the results of such studies, data will be available for numerical and theoretical model development and a better fundamental understanding of the phenomenon can be obtained.
- b) Numerical experiments should be conducted to model the behavior of dynamically loaded unsaturated soils using techniques such as the Distinct Element Method (DEM) in which load transfer mechanisms and material phase interactions can be examined. The modeling effort would be guided by the experimentally derived stress-strain data from SHPB and centrifuge tests.

2) A series of fully instrumented, carefully controlled small scale field explosive tests should be conducted in close coordination with additional laboratory studies. Test parameters should include variations in saturation, compaction methods, boundary conditions and applied energy. Field instrumentation should provide measurements of input energy, soil deformation and stress and transmitted energy as a minimum. Results will be useful for relating laboratory and field measured soil behavior, providing valuable input for material models.

REFERENCES

1. ASTM (1992) *Annual Book of ASTM Standards - Volume 04.08 Soil and Rock; Dimension Stone; Geosynthetics*, American Society for Testing and Materials, Philadelphia, PA, pp. 160-164.
2. Charlie, W.A. and Pierce, S.J. (1990) "High Intensity Stress Wave Propagation in Unsaturated Sands." *Report No. ESL-TR-90-12, Under Contract No. F49620-87-0004*, AFESC, Engineering Research Division, Tyndall AFB, FL, Sept., 146 p.
3. Charlie, W.A., Ross, C.A. and Pierce, S.J. (1988) "Split-Hopkinson Pressure Bar Testing of Unsaturated Sand." *Geotechnical Testing Jnl.*, GTJODJ, Vol. 13, No. 4, pp. 291-300.
4. Corey, A.T. (1977) *Mechanics of Heterogeneous Media*. Water Resources Publications, Ft. Collins, CO.
5. DeGregorio, V.B. (1990) "Loading Systems, Sample Preparation, and Liquefaction." *Jnl. of Geotechnical Engineering*, Vol. 116, No. 4, ASCE, pp. 805-821.
6. Farr, J.V. and Woods, R.D. (1988) "A Device for Evaluating One-Dimensional Compressive Loading Rate Effects." *Geotechnical Testing Jnl.*, GTJODJ, Vol. 11, No. 4, pp. 269-275.
7. Felice, C.W., Brown, J.A. Gaffney, E.S. and Olsen, J.M. (1987) "An Investigation into the High Strain-Rate Behavior of Compacted Sand Using the Split-Hopkinson Pressure Bar Technique." *Proc., 2nd Symp. on the Interaction of Non-Nuclear Munitions With Structures*, P.C. Beach, FL, pp. 391-396.
8. Felice, C.W., Gaffney, E.S., Brown, J.A. and Olsen, J.M. (1987) "Dynamic High Stress Experiments on Soils." *Geotechnical Testing Jnl.*, GTJODJ, Vol. 1, No. 4, pp. 192-202.
9. Fitzpatrick, B.J. and Veyera, G.E. (1992) "The Microstructure of Compacted Moist Sand and Its Effect on Stress Transmission." *Report No. ESL-TR-92-01 Under Contract Number F49620-88-C-0053/SB5881-0378*, AFESC/IRDCM, Tyndall AFB, FL, 200 p.
10. Friesenhahn, G.J. and Ross, C.A. (1985) "Transmission of Pressure Waves in Granular Materials Using a Split-Hopkinson Pressure Bar." *Response of Geologic Materials to Blast Loadings*, ASCE/ASME Mechanics Conf., Albuquerque, NM, pp. 143-147.
11. Gaffney, E.S., Brown, J.A. and Felice, C.W. (1985) "Soil Samples for the Split Hopkinson Bar." *Proc., 2nd Symp. on the Interaction of Non-Nuclear Munitions With Structures*, P.C. Beach, FL, pp. 397-402.

12. Hendron, A.J. (1963) "The Behavior of Sand in One-Dimensional Compression." *Doctoral thesis submitted to the University of Illinois at Urbana, IL, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.*
13. Hendron, A.J., Davisson, M.T. and Paroloa, J.F. (1969) "Effect of Degree of Saturation on Compressibility of Soils From the Defence Research Establishment, Suffield." *Report No. WES-BPJ-68-67*, U.S. Army Waterways Experiment Station, Vicksburg, MS.
13. Jackson, J.G., Ehrgott, J.Q. and Rohani, B. (1981) "Loading Rate Effects on Compressibility of Sand." *Jnl. of the Geotechnical Engineering Division*, ASCE, Vol. 106, No. GT8, August, pp. 839-852.
14. Juang, C.H. and Holtz, R.D. (1986) "Fabric, Pore Size Distribution, and Permeability of Sandy Soils." *Jnl. of Geotechnical Engineering*, Vol. 112, No. GT9, ASCE, pp. 855-868.
15. Ladd, R.S. (1977) "Specimen Preparation and Cyclic Stability of Sands." *Jnl. of the Geotechnical Engineering Division*, ASCE, Vol. 103, No. GT6, pp. 535-547.
16. Lindholm, U.S. (1971) "High Strain Rate Tests." *Measurement of Mechanical Properties, Vol. V Part 1*, Interscience Publishers, New York, NY, pp. 200-271.
17. Mahmood, A. and Mitchell, J.K. (1974) "Fabric-Property Relationships in Fine-Grained Materials." *Clays and Clay Minerals*, Vol. 22, pp. 397-408.
18. Mitchell, J.K. (1976a) *Fundamentals of Soil Behavior*, John Wiley and Sons, New York, NY.
19. Mitchell, J.K. (1976b) "The Influences of Sand Fabric on Liquefaction Behavior." *Report Number S-76-5*, U.S. Army Waterways Experiment Station, Vicksburg, MS, pp. 1-38.
20. Mulilis, J.P., Seed, H.B., Chan, C.K., Mitchell, J.K. and Arulanandan, K. (1977) "Effects of Sample Preparation Technique on Sand Liquefaction." *Jnl. of the Geotechnical Engineering Division*, ASCE, Vol. 103, No. GT2, pp. 91-108.
21. Oda, M. (1972a) "Initial Fabrics and Their Relations to Mechanical Properties of Granular Material." *Japanese Society of Soil Mechanics and Fndtn. Engineers*, Tokyo, Japan, Vol. 12, No. 1, pp. 17-36.
22. Oda, M. (1972b) "The Mechanism of Fabric Changes During Compressional Deformation of Sand." *Japanese Society of Soil Mechanics and Fndtn. Engineers*, Tokyo, Japan, Vol. 12, No. 2, pp. 1-18.
23. Richart, F.E., Jr., Hall, J.R. and Woods, R.D. (1970) *Vibrations of Soils and Foundations*, Prentice-Hall, Inc, Englewood Cliffs, NJ.
24. Ross, C.A. (1989) "Split-Hopkinson Pressure Bar Tests." *Report No. ESL-TR-88-2*, AFESC/RDCM, Tyndall AFB, FL, 80 p.
25. Ross, C.A., Nash, P.T. and Friesenhahn, C.J. (1986) "Pressure Waves in Soils Using a Split-Hopkinson Pressure Bar." *Report No. ESL-TR-86-29*, AFESC, Tyndall AFB, FL, July, 83 p.
26. Veyera, G. E. (1989) "Static and Dynamic Behavior of Compacted Unsaturated Sands." *Final Report to US AFOSR Under Contract No. F49620-87-0004*, AFESC/RDCM, Tyndall AFB, FL, Sept., pp. 45-1:45-20.
27. Veyera, G.E. and Fitzpatrick, B.J. (1990) "A Specimen Preparation Technique for Microstructural Analysis of Unsaturated Soils." *Final Report to US AFOSR Under Contract No. F49620-88-C-0053*, Bolling AFB, Washington D.C., Sept., 20 p.
28. Wilbeck, J.S., Thompson, P.Y. and Ross, C.A. (1985) "Laboratory Measurement of Wave Propagation in Soils." *Proc. of the 2nd Symp. on the Interaction of Non-Nuclear Munitions With Structures*, P. C. Beach, FL, pp. 460-465.
29. Whitman, R.V. (1970) "The Response of Soils to Dynamic Loadings; Report 26, Final Report." *Contract Report No 3-26*, US Army Waterways Experiment Station, Vicksburg, MS.
30. Whitman, R.V., Miller, E.T. and Moore, P.J. (1964) "Yielding and Locking of Confined Sand." *Jnl. of the Soil Mechanics and Fndtn. Division*, ASCE, Vol. 90, No. SM4, April, pp. 57-84.
31. Wu, S., Gray, D.H. and Richart, F.E. (1984) "Capillary Effects on Dynamic Modulus of Sands and Silts." *Jnl. of the Geotechnical Engineering Division*, ASCE, Vol. 110, No. GT9, Sept., pp. 1188-1202.

APPLICATION OF FIBER-OPTIC LASER FLUORESCENCE SPECTROSCOPY TO
ENVIRONMENTAL MONITORING

Brian S. Vogt
Professor
Department of Chemistry
Bob Jones University
Greenville, SC 29614

Final Report for:
Summer Research Program
Air Force Civil Engineering Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

APPLICATION OF FIBER-OPTIC LASER FLUORESCENCE SPECTROSCOPY TO
ENVIRONMENTAL MONITORING

Brian S. Vogt
Professor
Department of Chemistry
Bob Jones University
Greenville, SC 29614

Abstract

Recent acquisition of a fiber-optic laser fluorescence spectrometer in the Environics Division of the Headquarters Air Force Civil Engineering Support Agency (HQ AFCEA/RAV) has opened up possibilities for rapid characterization of model aquifers and contaminated Air Force Sites. The goals of this project were to familiarize research personnel with the operation of the spectrometer, characterize spectrometer response, incorporate new computer and optical components into the system, optimize fiber-optic probe design, employ the instrument in a model aquifer study, and document the principles underlying the instrument and its application for distribution to Air Force personnel not familiar with the system.

All but two of the goals were achieved. Problems with laser cooling were diagnosed and resolved. Spectrometer response to naphthalene was characterized over a range of conditions. Several modifications of fiber-optic probe design were considered and tested. A prototype of the most promising design was designed, constructed, and tested. A refined and scaled-down version of the prototype was designed, constructed, tested, and found to be highly successful in monitoring not only analyte fluorescence, but also light scattered by turbidity introduced into samples. Results indicate that it may be possible to correct fluorescence signals for the effects of turbidity. Difficulties with procurement prevented the incorporation of additional computer and optical components into the system. There was not enough time to employ the spectrometer in a model aquifer study.

Recommendations were made concerning refinements in probe design and construction, further studies dealing with turbidity, studies dealing with potential interferences, and the application of the system to model aquifer studies. Additional recommendations were made as to how to accelerate the development and transfer of this technology to the field.

APPLICATION OF FIBER-OPTIC LASER FLUORESCENCE SPECTROSCOPY TO ENVIRONMENTAL MONITORING

Brian S. Vogt

INTRODUCTION

The Environics Division of the Headquarters Air Force Civil Engineering Support Agency (HQ AFCESA/RAV) is aggressively exploring and applying new technology to the problems of monitoring and remediating environmental contamination at Air Force sites throughout the world. Ongoing research into the fate and transport of ground-water contaminants is being carried out using model aquifers. Mathematical modeling of the results gives greater insight into how contaminant plumes are distributed throughout spill sites and thus contributes to more cost-effective site characterization and cleanup. Many such studies have been performed using chromatography and/or liquid scintillation counting to analyze samples from model aquifers. Unfortunately, these methods require the removal and preparation of samples prior to analysis, resulting in relatively long analysis times for each sample.

A fiber-optic laser fluorescence spectrometer was installed at HQ AFCESA/RAV in May, 1992. This system was designed and assembled by G.D. Gillipsie and his research group [1]. A fiber-optic probe coupled to the laser can be placed into a small-scale monitoring well in a model aquifer. Fluorescence measurements acquired in only 10 to 60 seconds are sufficient to obtain the concentration of a given contaminant without removing a sample for analysis. This not only enables faster data acquisition, but also more thorough characterization of a given system.

A transportable version of this spectrometer will allow rapid, cost-effective analysis of contaminants in existing monitoring wells at Air Force Sites. Laser spectrometer field analysis of toluene contamination in wells at Tinker Air Force Base (Oklahoma) was highly encouraging [2].

Use of a transportable laser spectrometer system in conjunction with a cone penetrometer provides further cost savings. A cone penetrometer is a truck-mounted system that hydraulically pushes a hardened steel tube into the ground at a rate of one meter per minute. The tube contains a variety of sensors or soil and groundwater sampling tools. Cone penetrometry allows for the determination of soil type and stratigraphy with little or no prior site preparation. Including a fiber-optic probe in the penetrometer tube permits rapid spectroscopic

characterization of the site without drilling monitoring wells.

GOALS

The long-term application of the HQ AFCEA/RAV laser spectrometer is not only to facilitate model aquifer studies, but also to enable the development and refinement of analytical procedures to be applied in the characterization of contaminated Air Force sites. Short-term goals for the Summer of 1992 were as follows:

1. Familiarize research personnel with the operation of the spectrometer.
2. Characterize instrumental response.
3. Incorporate new computer and optical components into the system, install the relevant software, and phase them into use.
4. Optimize fiber-optic probe design.
5. Employ the instrument in a model aquifer study.
6. Document the principles underlying the instrument and its application for distribution to Air Force personnel not familiar with the system.

APPARATUS

Figure 1 shows the layout of the laser spectrometer. The primary laser is a 20 Hz Surelite Series 20 Nd:YAG laser (Continuum, 3150 Central Expressway, Santa Clara, CA 95051). The 1064 nm fundamental is frequency quadrupled to give 266 nm output. The dye laser illustrated will be installed in the future. A portion of the collimated 266 nm output is directed through a launch fiber (600 μ m core diameter Superguide fiber from Fiberguide Industries, 1 Bay St., Stirling, NJ 07980) to excite fluorescence in the sample to be analyzed. A fraction of the fluorescence is directed by a collection fiber to a monochromator. Two-fiber probes were used because current application of this system requires that probes be as small as possible (0.25 inch or less in outside diameter). Light selected by the monochromator is sent to a photomultiplier tube powered by a Model PS325 digital power supply (Stanford Research Systems, Inc., 1290 D Reamwood Ave., Sunnyvale, CA 94089). The monochromator used this summer was on loan from North Dakota State University [1] and will be replaced by a computer-controlled monochromator in the near future. Output from the photomultiplier tube

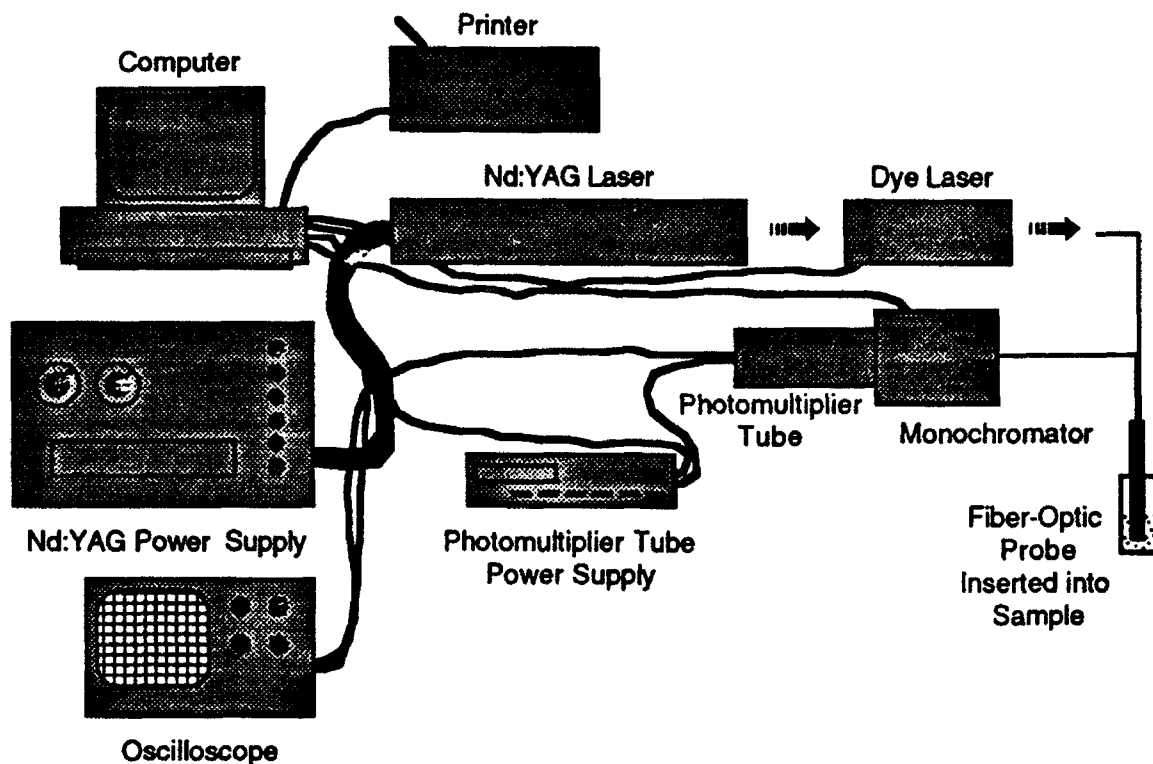


Figure 1. Envirionics Fiber-Optic Laser Fluorescence Spectrometer

is sent to a Model 2440 digital storage oscilloscope (Tektronix, Inc., P.O. Box 500, Beaverton, OR 97077) for digitization and display. Various waveform parameters can be calculated and displayed by this oscilloscope. The photomultiplier tube power supply, oscilloscope, dye laser, and new monochromator will be controlled remotely by the computer. This will permit data acquisition and analysis to be performed in a semi-automated fashion.

RESULTS

One of the Summer goals was to document the principles underlying the laser spectrometer system and its application to environmental monitoring. A paper [3] has been written for distribution to Air Force personnel interested in learning more about the instrument.

Severe drift in laser power was encountered early in the summer. Because a laser power meter was not available, laser output was monitored by measuring the intensity of the water Raman signal at 293 nm as a function of time. It was found that this signal would drop with time but revert to its initial value after the laser was allowed to cool down.

The laser system operates with two cooling loops. The internal loop is a flow of cooling water from the laser power supply to the laser head. Tap water flows through the the external loop, cools the water in the internal loop, and goes down the drain. It was found that the tap water temperature in the laboratory reached 85° F, far too warm to satisfy cooling requirements. Tap water flow was replaced with a flow of ethylene glycol circulated through a Model HX-50 constant-temperature bath (Neslab Instruments, Inc., Newington, NH, 03801) thermostatted at 20°C. This stabilized laser output for a short period of time, after which it again degraded. Partial disassembly of the internal cooling loop revealed that a screen mesh filter was substantially blocked with algae deposits. In keeping with the manufacturer's procedure, the entire internal cooling system was cleaned with hydrogen peroxide and flushed with distilled water. Subsequent output was reasonably stable, as is illustrated in Figure 2.

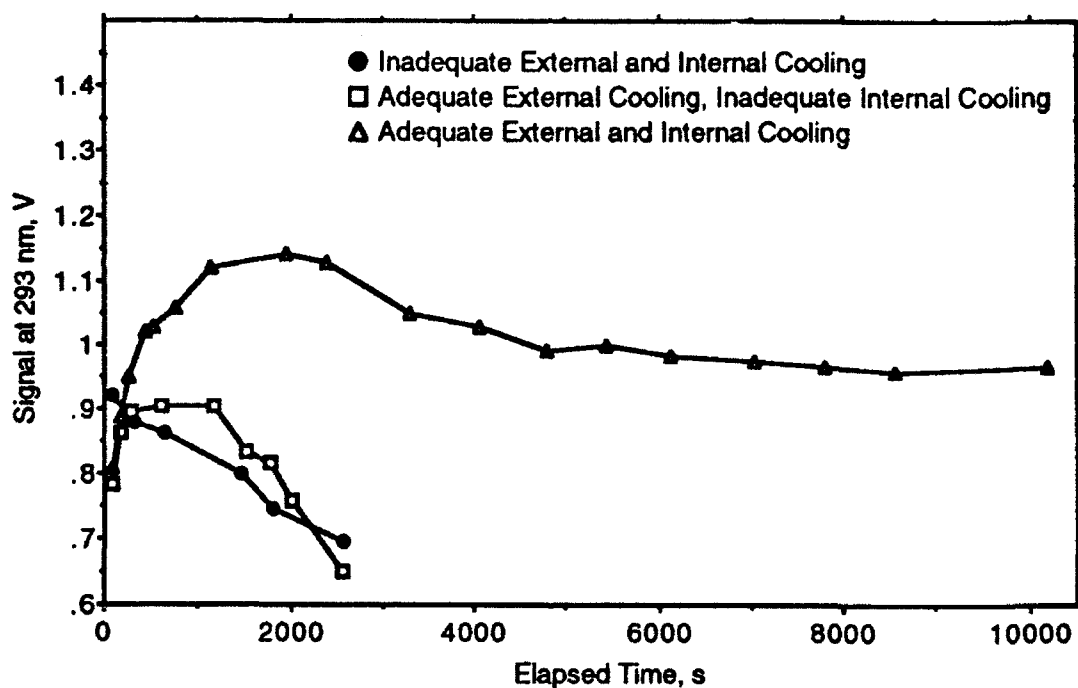


Figure 2. Laser Performance with Different Levels of Cooling.

Naphthalene was the primary organic investigated in this work because naphthalene is a component of jet fuel (JP-4, JP-8, etc.) and is often found in the ground-water at contaminated Air Force sites.

Once a naphthalene decay profile was acquired and displayed, the oscilloscope easily calculated peak height, peak area, and many other parameters. It seemed desirable to use peak area as an indication of

naphthalene concentration. Unfortunately, a quirk in the scope made it necessary to perform some inconvenient and time-consuming interpolation to obtain accurate peak areas. Peak-to-peak heights, which are the signals at the maxima of the fluorescence decay profiles, and accurate peak areas were obtained for a series of naphthalene solutions ranging in concentration from 2 ppb to 20 ppm. Figure 3 is a plot of peak height as a function of peak area. Each point, therefore, represents a peak height and area for a single naphthalene solution. As can be seen, the peak heights and areas correlated extremely well, permitting us to use easily obtainable peak heights instead of peak areas.

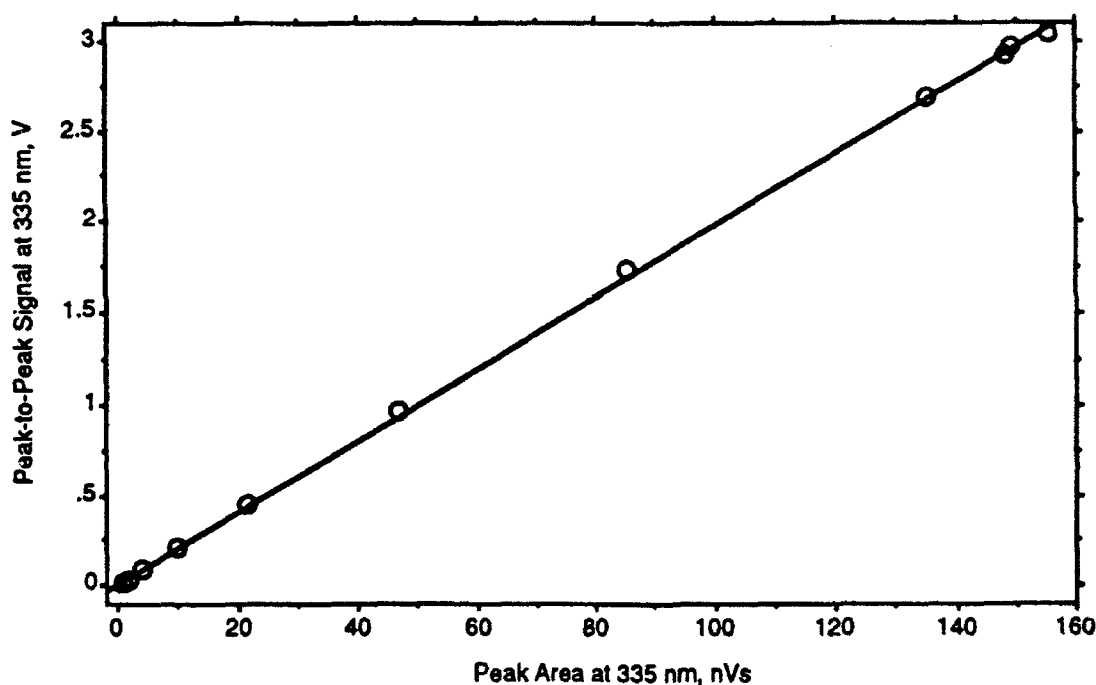


Figure 3. Correlation Between Peak Height and Peak Area.

It was also questioned whether signals at points later than the maximum of the naphthalene fluorescence decay profile would prove as useful. Signals were acquired at the maximum of the decay profile as well as at 12 ns, 25 ns, and 40 ns past the maximum. Figure 4 shows calibration curves constructed from these data. Although signals at any of these times can be used, the peak-to-peak height at the maximum is clearly the most sensitive measure of naphthalene concentration. New preliminary data, however, suggest that one of the later times might be more suitable if turbidity in the sample scatters excessive laser excitation light into the collection fiber [4].

Data were signal-averaged with the oscilloscope. The effect of

signal acquisition time was investigated by measuring signal as a function of time for several naphthalene solutions. Triplicate runs for a 0.0201 ppm naphthalene solution were performed and averaged-the results

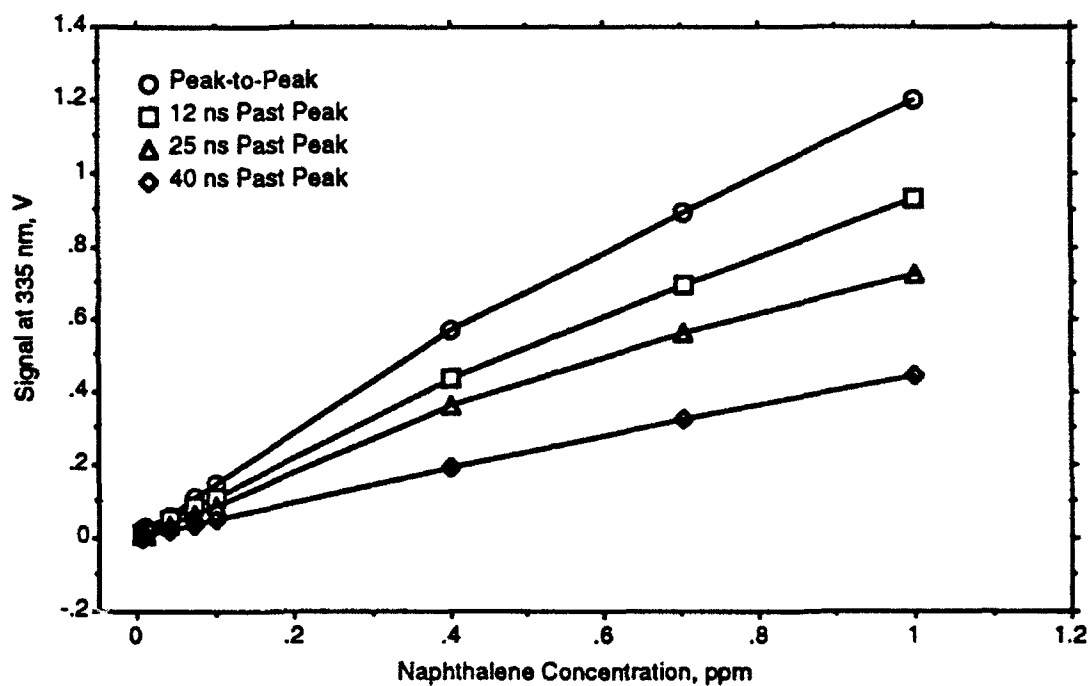


Figure 4. Data Obtained at Different Points on the Decay Profile.

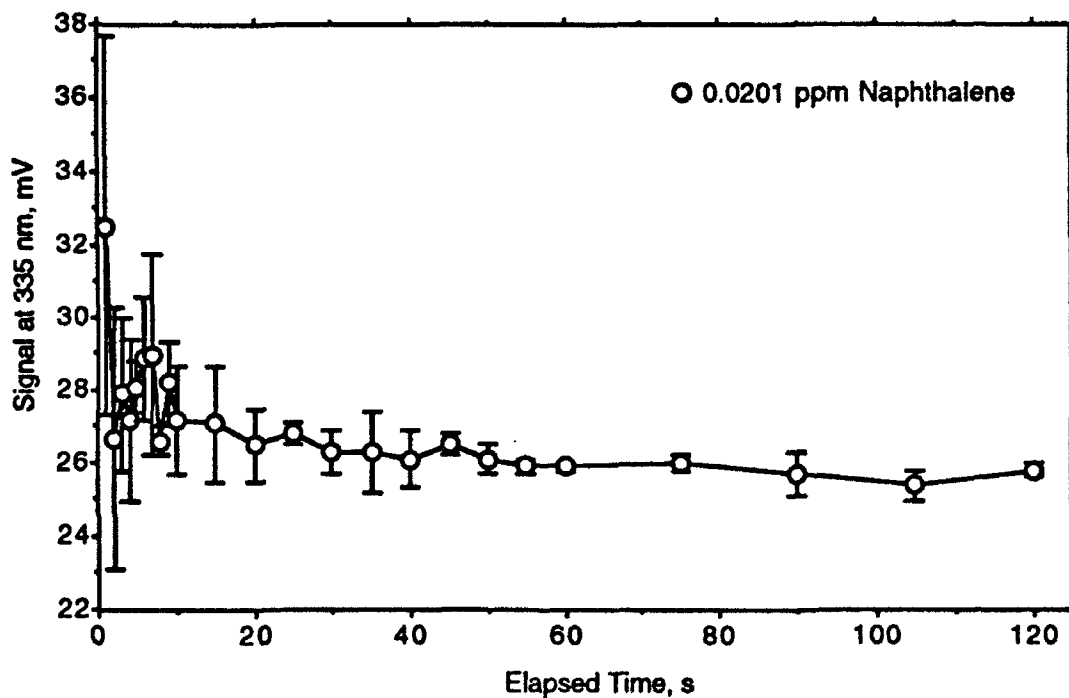


Figure 5. Signal versus Data Acquisition Time.

are shown in Figure 5. In this case it is clear that data acquisition times of about 20 seconds or longer give the best data. Consideration of other data suggested that 30 second acquisition times would be the most prudent. However, these data were acquired before the inadequate laser cooling situation was remedied. These measurements need to be repeated now that laser output is stable.

Some of the signals measured were sensitive to the position of the probe in the sample vessel. It seemed likely that this was due to the reflection of light off the walls of the container. This would adversely affect measurements in wells in model aquifers.

Furthermore, turbidity is often encountered in the water contained in monitoring wells in both model aquifers and field sites. Turbidity can scatter both excitation light and fluorescence and thus alter fluorescence decay profiles.

Several efforts were made to design fiber-optic probes that would not be sensitive to reflection. Figure 6 shows the designs that were considered and tested. Laser light reflected into the collection fiber in design ① was so intense that the design had to be abandoned. Design ② may have worked had it been possible to put a sharp enough point on the teflon rod. Design ③ showed some promise in that it did cut down on the amount of laser light that entered the collection fiber. However, it would have been impractical to mount and position the beveled brass rod on the end of the fibers. Design ④ is a variation of design ③ that permits easy construction and alignment. Initial results were so encouraging that design ④ was refined to permit construction of a more durable probe. An aluminum prototype 1/2" in diameter was designed, built, and tested before constructing a 1/4" stainless steel probe. Details of the construction of this probe, hereafter referred to as the "new probe," are shown in Figure 7. The term "old probe" will be used to refer to the simple two-fiber probe without the modifications depicted in Figure 6. It will be seen that the geometry of the open section is designed not only to give the fibers access to the solution, but also to direct the path of nonabsorbed excitation light.

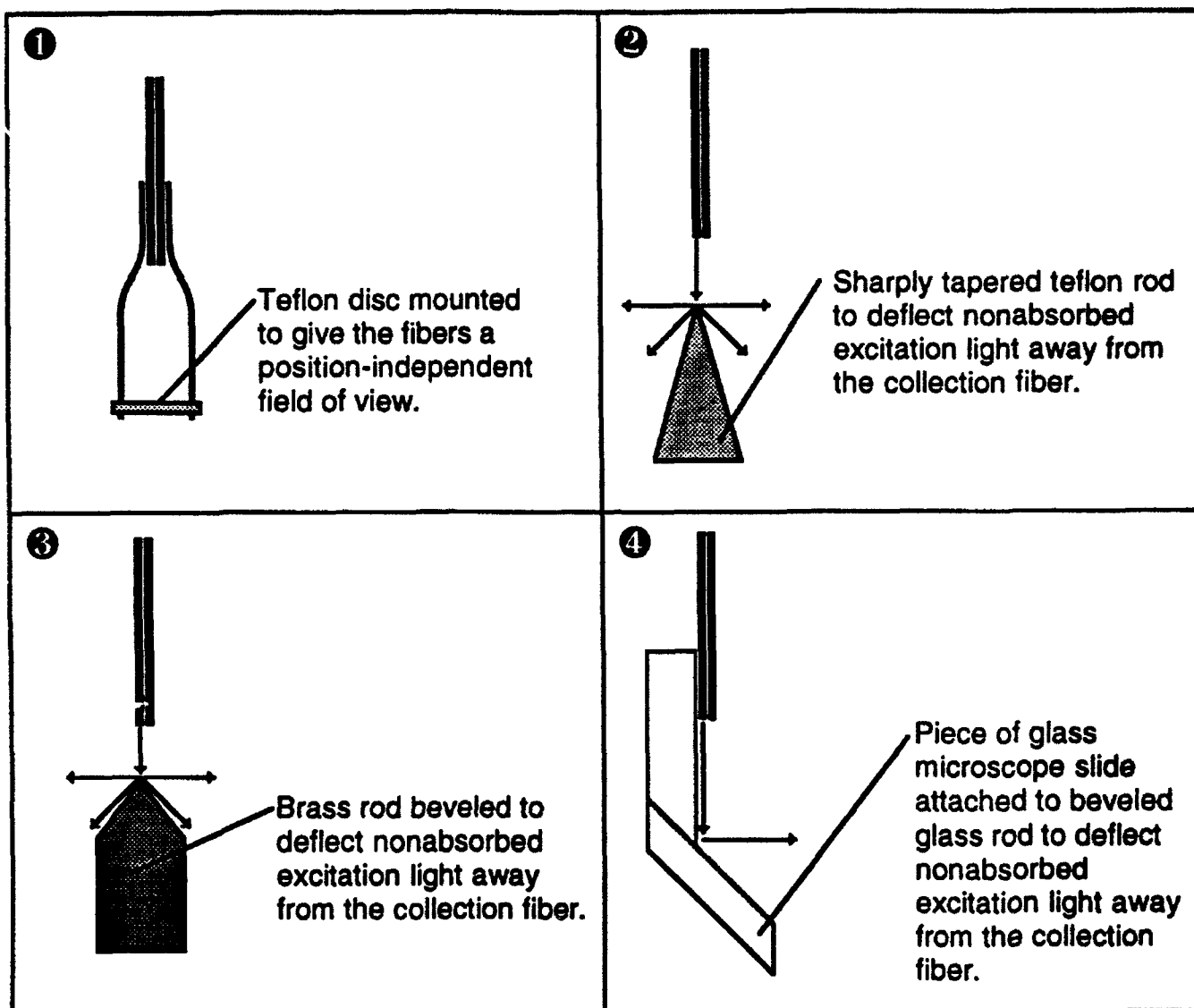


Figure 6. Probe Designs.

The detail in Figure 7 shows that the fields of view of the two fibers overlap. According to the manufacturer, the numerical aperture of the fiber in air is 0.4, from which it is easily calculated that the cone of acceptance is about 47° . In water, however, the refractive index is 1.333 [5], and thus the cone of acceptance in water is about 35° . The highly collimated laser light enters the launch fiber parallel to the long axis of the fiber. This results in far lower dispersion of the light coming out of the launch fiber than would be observed if the light were noncollimated or if it entered at an off-axis angle. Laboratory measurements indicated approximately 15° and 11° cones of laser light exiting the launch fiber in air and water, respectively.

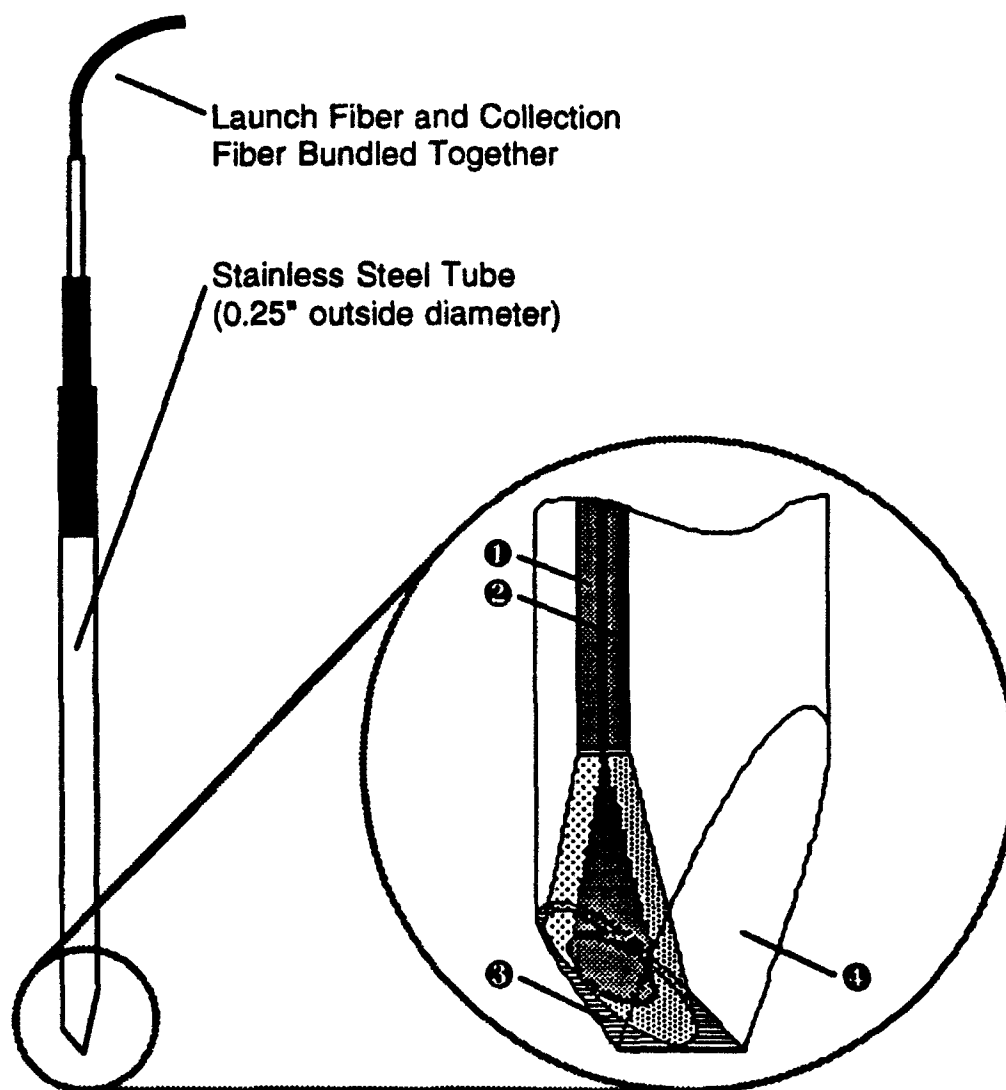


Figure 7. New Probe Construction.

The fact that the excitation light leaves the launch fiber in an 11° cone but the collection fiber views a 35° cone is illustrated in Figure 8, which is a more accurate and detailed illustration of the light path through the new probe.

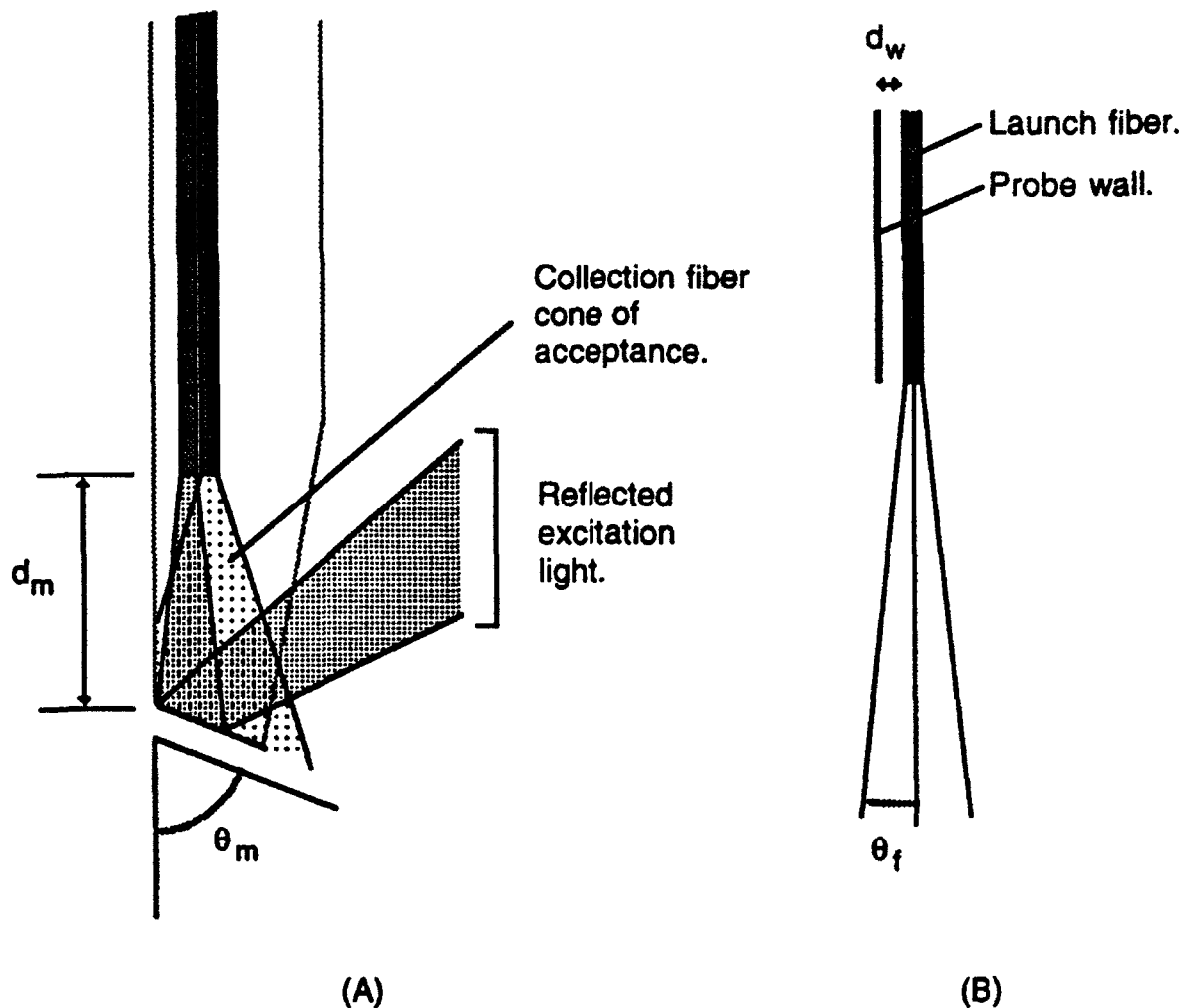


Figure 8. Important Parameters and Light Path in the New Probe.

A good way of obtaining square, smooth fiber ends is to embed the fiber end in epoxy and cut and polish the end after the epoxy is dry. This procedure could not be followed because the polishing kit on order did not arrive on time. Consequently, all of the fibers used in this study were hand cleaved by scoring and snapping the fiber, which results in imperfections on the fiber core surface. This procedure also invariably stretches and tears some of the nylon buffer around the fiber at the point of cleavage (this is easy to see under a microscope). Both of these factors may influence the shapes of the light cones, which should be checked when new probes are constructed with polished fiber ends instead of cleaved ends.

The launch fiber must be placed some distance from the probe wall to ensure that all of the excitation light hits the glass mirror and that none of it strikes the probe wall. Excitation light striking the probe wall would be scattered and probably introduce an undesired background signal. The distance between the launch fiber and probe is designated d_w and is illustrated in Figure 8 (B). Even with the launch fiber placed properly relative to the probe wall, it must be close enough to the mirror to further avoid scattering at the wall. To accomplish this the tip of the launch fiber cannot be placed any further away from the mirror than d_m , which is illustrated in Figure 8 (A).

In this research it was found convenient to construct the probe with a spacer of known thickness placed between the probe wall and the launch fiber. This fixes d_w at a known value and permits easy calculation of d_m . Basic geometry shows that $\tan\theta_f = d_w/d_m$, where θ_f is the half angle of the light exiting the launch fiber-see Figure 8 (B). The launch and collection fibers were epoxied together before positioning them in the steel body of the probe. Before attaching the fibers to the steel tube, the tube was beveled at the mirror angle, θ_m . This angle is illustrated in Figure 8 (A). To be on the safe side, the mirror must be attached at an angle that guarantees that none of the excitation light will be reflected directly into the collection fiber. Additional geometry shows that this condition is met when $\theta_m < (90 - \theta_f)$. The probe used in this study had $\theta_f = 5.5^\circ$, $d_w = 0.7$ mm, $d_m = 7$ mm, and $\theta_m = 60^\circ$. The orientation of the fibers relative to the glass mirror and probe body can be discerned by looking at the ellipses of light projected onto the mirror in Figure 7. After the fibers were fixed in position, the mirror was epoxied in place. With the mirror firmly attached, the open section was cut by grinding on a wet grinder equipped with a fine diamond wheel. Excess glass at the edges of the mirror was ground off with the same wheel. Heat shrink was used to protect the fibers where necessary and also to connect the two fibers together to make a single cable. The heat shrink was applied carefully to avoid melting the nylon buffer.

Figure 8 illustrates the angle followed by nonabsorbed excitation light that is reflected off the glass mirror. The glass mirror is angled so that any nonabsorbed excitation light exits the probe at an upward angle to prevent it from being reflected back into the probe by the walls of the sample container. Furthermore, measurements indicated that the glass used absorbs 266 nm light efficiently [4]. Consequently, any excitation light that enters into the glass mirror (instead of being reflected off the surface) is not transmitted into the sample below the probe and as such has no opportunity to be reflected back up into

collection fiber. Notice that excitation light passing out of the probe passes through the field of view of the collection fiber. This is beneficial because it can excite additional fluorescence in front of the collection fiber and thereby improve the signal. This arrangement is highly effective: measurements of naphthalene fluorescence at 335 nm in water indicated that most of the background problem experienced with the old probe was gone. However, even this arrangement is not perfect: tuning the monochromator to 266 nm revealed that excitation light was still picked up by the collection fiber. It is not known if this is due to scattering within the probe or crosstalk between the fibers. Crosstalk could be easily eliminated by painting or putting UV-absorbing heat shrink around one or both of the fibers.

It was also found that fluorescence measurements with the new probe were essentially independent of the probe position in the sample container. This is attributed partly to the path of nonabsorbed excitation light, as discussed above. Another reason may be that much of the view of the collection fiber is the controlled environment inside the probe. There is a fairly high probability that any light that does bounce around inside the sample container will be deflected away from the collection fiber by the steel probe jacket and the glass mirror.

Turbidity can be quantified by measuring the amount of light transmitted through a solution (turbidimetry) or the amount of light scattered by a solution (nephelometry). The latter is easily done not only in a conventional fluorimeter, but also with the new probe. Figure 9 shows the response of the new probe to varying concentrations of clay (Ca-Montmorillonite, "Cheto," from Apache County, AZ) obtained through the Clay Minerals Society/Source Clay Minerals Repository (Department of Geology, University of Missouri, Columbia, MO 65211).

The response to the laser light is encouraging. It had been suggested that the Raman scatter of water at 293 nm might be sensitive to turbidity, but Figure 9 shows that this is not so.

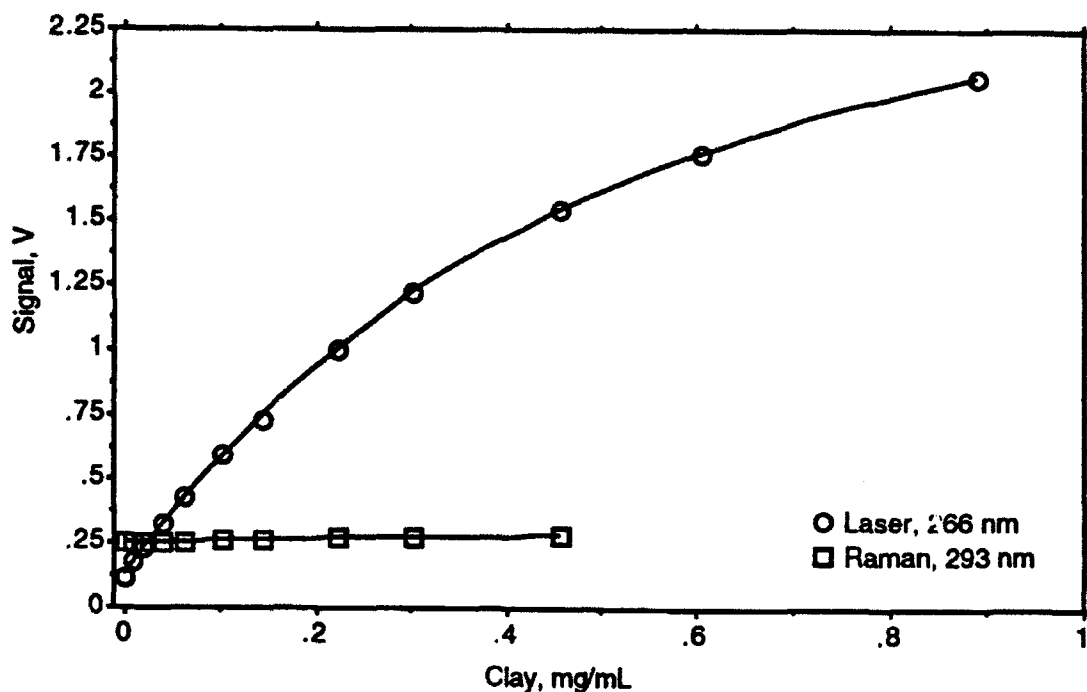


Figure 9. Response of the New Probe to Turbidity.

Figure 10 shows that the linear range of naphthalene calibration data acquired with the old probe extends to about 0.1 ppm. It was observed that the plot curved downward at concentrations higher than shown in Figure 10: this behavior is commonly seen in fluorescence calibration data. Each point is the average of four separate runs of the appropriate concentration. Different concentrations correspond to different naphthalene solutions in different volumetric flasks. The old probe was inserted about 1 cm into the liquid before taking fluorescence measurements.

There is scatter around the fitted line even though the relative standard deviations were comparatively low. This may have been due to the fact that different flasks had different heights and internal geometries, which in turn could have reflected varying amounts of light back to the collection fiber.

Similar data acquired with the new probe are shown in Figure 11 [4]. These data also curve downward at concentrations higher than shown in the Figure. In contrast to the old probe, the linear range of the new probe extends to approximately 0.4 ppm naphthalene.

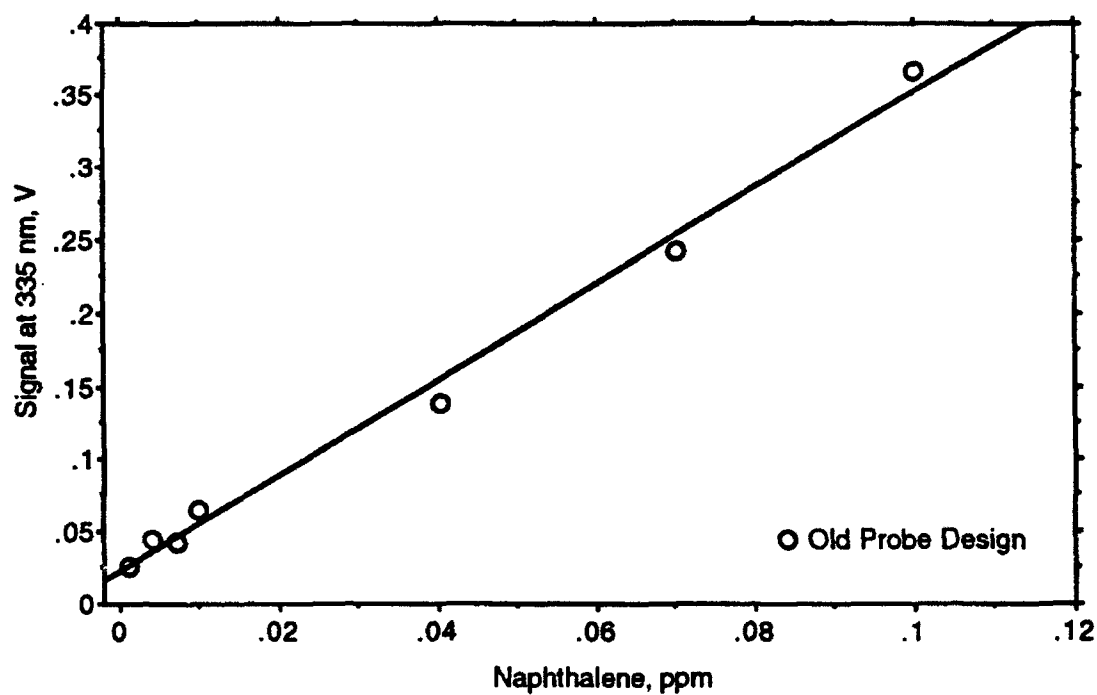


Figure 10. Naphthalene Calibration Data: Old Probe

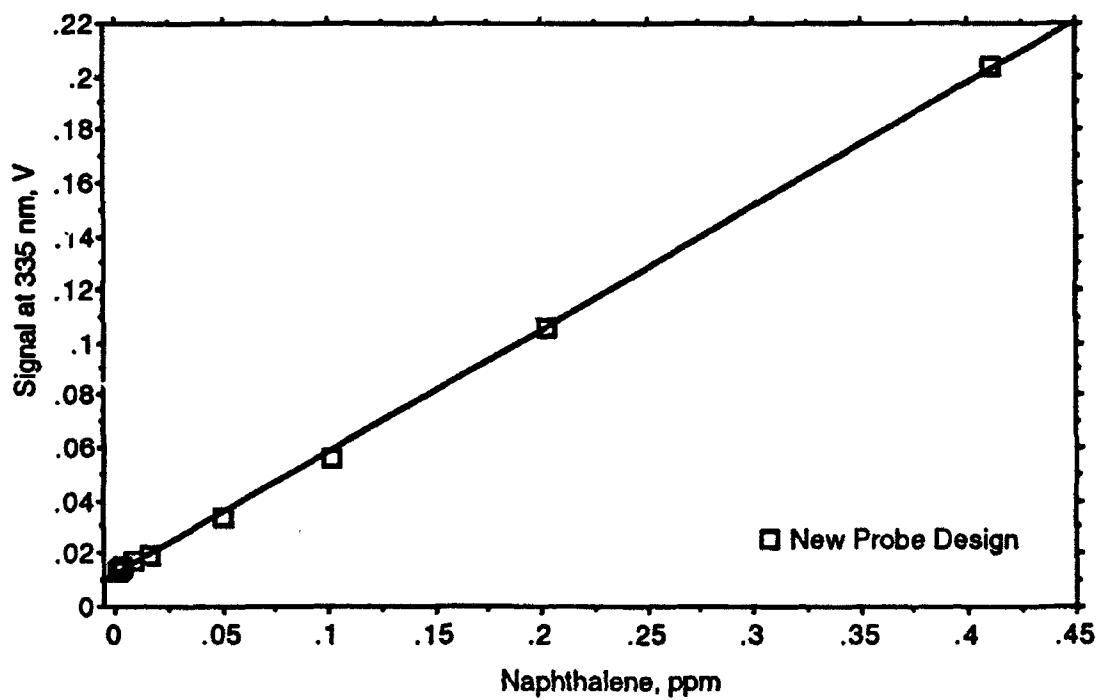


Figure 11. Naphthalene Calibration Data: New Probe

Comparison of the new probe design to designs in the literature reveals some interesting differences. A fiber-optic probe has been designed and used successfully to perform particle sizing by measuring the scatter of white light [6]. However, the geometrical design of this probe is not amenable to being scaled down to meet Environics size requirements. An apparatus consisting of fibers placed in front of various reflective surfaces was constructed and tested [7]. The fibers were perpendicular to the surfaces. Not surprisingly, it was found that signals were increased due to reflection. The fibers and reflective surfaces in this apparatus were not fabricated into an integrated probe. A simple probe was constructed and tested with suspensions of latex spheres [8]. One end of a short stainless steel rod was bored out so that it looked much like a thimble. The fibers were inserted into a hole in the other end of the rod so that their acceptance cones projected into the hollow portion of the rod. This probe was applied to theoretical studies of light scattering but not to fluorescence or turbidity measurements.

It is felt that the probe designed in this research compares favorably to those in the literature and that it offers a unique combination of size and capability.

All but two of the Summer goals were reached. It was not possible to incorporate new electronic or optical components into the laser system because of procurement difficulties. Application of the laser spectrometer to a model aquifer study might have begun had it not been for the significant amount of time spent on diagnosing and rectifying problems with laser cooling.

RECOMMENDATIONS

Probe design needs to be refined to give more durable probes. The current design probably would not stand up under use in model or field wells because the glass mirror is exposed. This could probably be remedied by adding a steel sleeve around the bottom of the probe and filling the space below the mirror with epoxy. This is illustrated in Figure 12.

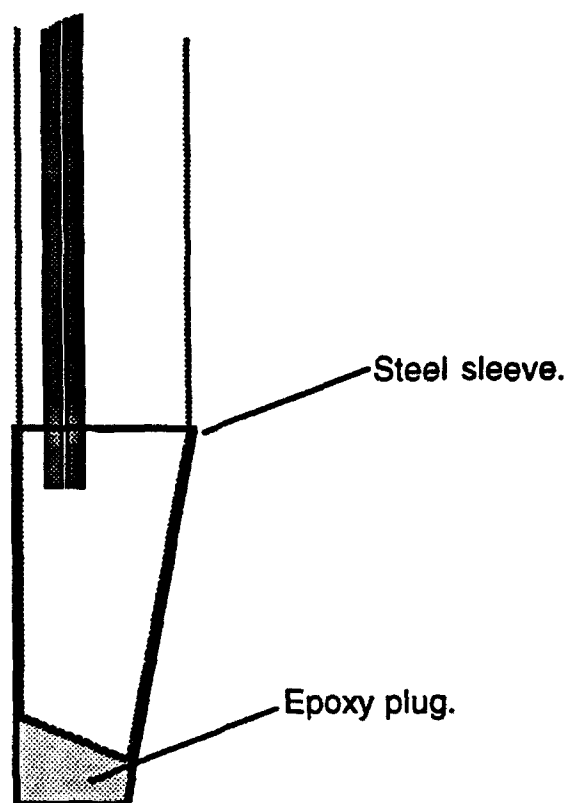


Figure 12. Probe Modification to Increase Durability.

Work needs to be done to see if the probe can be scaled down from an outside diameter of $1/4"$ to $1/8"$. This will probably require changes in probe construction techniques.

Use of the new probe to measure turbidity has been encouraging. However, further research needs to be done to see how the response to different scattering agents (sand, titanium dioxide, etc.) compares. Furthermore, experiments with different excitation wavelengths should be

performed once the dye laser is incorporated into the instrument. Data show that naphthalene fluorescence decreases with increasing turbidity [4]. The relationship between fluorescence signals and turbidity needs to be determined so that turbidity measurements can be used to correct fluorescence signals.

Quenching, energy transfer, exciplex formation, and excimer formation are processes that can affect the fluorescence spectra and lifetimes of fluorescent substances. Whether these processes are significant or not depends on the fluorescence lifetimes and concentrations of the species involved. Unfortunately, sample matrices at contaminated Air Force sites have the potential of being very complex. Jet fuel contains many components. Various types of solvents may have been spilled. It would be useful to see how realistic concentrations of potential interferences affect the emission properties of important contaminants such as benzene, toluene, naphthalene, etc., in simple solutions and in complex matrices such as jet fuel.

Model aquifer studies should be performed not only to model the transport of contaminants in the environment, but also to obtain valuable information about the potential performance of the system in the field.

If adequate funds become available, several things could be done to accelerate the development and application of this technology. A second laser system could be acquired and dedicated to the development of analytical fluorescence procedures. If this is done, the current system could be dedicated to model aquifer studies. A third system could be acquired and dedicated to the development of analytical Raman procedures. Additional personnel would have to be brought into Environics to operate the additional laser systems. Periodic field demonstrations would give a realistic assessment of where the technology is at. Taken as a whole, doing all of this development work simultaneously would very likely get the technology to the field in one-third to one-half the time it will take at the current rate of development.

REFERENCES

1. G.D. Gillispie, North Dakota State University, Department of Chemistry, Ladd Hall, Fargo, ND 58105-5516.
2. G.D. Gillispie and R.W. St. Germain, "In-Situ Tunable Laser Fluorescence Analysis of Hydrocarbons," Environmental and Process Monitoring Technologies, Tuan Vo-Dinh, Editor, Proc. SPIE 1637, pp. 151-162 (1992).
3. B.S. Vogt, "Principles of Fiber-Optic Laser Fluorescence Spectroscopy and Its Application to Environmental Monitoring," filed at HQ AFCESA/RAV, Tyndall Air Force Base, FL 32403-6001 (August, 1992).
4. T. Tipton (HQ AFCESA/RAV, Tyndall Air Force Base, FL 32403-6001), personal communication. Dr. Tipton acquired the data referred to.
5. "Optics Guide 5," Melles Griot, Inc. (1770 Kettering St., Irvine, CA 92714), p. 1-9 (1990).
6. Y. Kitagawa, A. Hayashi, and S. Minami, "Fiber-Optic Particle Size Monitor Based on White-Light Scattering," Applied Optics, 31(7), pp. 859-865 (1992).
7. C. Komives and J.S. Schultz, "Fiber-Optic Fluorometer Signal Enhancement and Application to Biosensor Design," Talanta, 39(4), pp. 429-441 (1992).
8. H.S. Dhadwal, C. Wu, and B. Chu, "Fiber Optic Detector Probes for Laser Light Scattering," Applied Optics, 28(19), pp. 4199-4205 (1989).

APPLICATIONS OF SMES IN AIR FORCE

**Xingwu Wang
Assistant Professor
Department of Electrical Engineering**

**Alfred University
Alfred, NY 14802**

**Final Report for:
Summer Research Program
Civil Engineering Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

APPLICATIONS OF SMES IN AIR FORCE

Xingwu Wang
Assistant Professor
Department of Electrical Engineering
Alfred University

Abstract

The concept of Superconducting Magnetic Energy Storage (SMES) had been examined by government agencies and utility industries for several decades. However, in the public domain, we could not find publications dealing with the applications of SMES in Air Force. In the past eight weeks, we carried out a technical study. Besides an extensive literature search, we discussed various issues with the leading experts on SMES, and talked to personnel related to electricity usage in different Air Force installations. Our findings are summarized in this report. In section I, development of SMES is briefly introduced. In section II, operational principles of SMES is described. In section III, seven possible areas of the applications are examined: (1) regular base, (2) bare base, (3) communications squadron, (4) air traffic control, (5) Sector Operational Control Center (SOCC) Radar stations, and (6) energetic weapon. Other six possible areas are also suggested for future studies, including space command and intelligence command. In section IV, discussions and recommendations are presented. The recommendations include both short and long term plans. In the Air Force applications, SMES can be used as an energy storage unit, or a power conditioning unit. The required energy of SMES can be divided into three ranges: 20 MW-Hr (72 GJ), 100-500 MJ, and 1-20 MJ. The 20 MW-Hr unit is stationary, and similar to that of an Engineering Test Model (ETM) currently being studied by the Defense Nuclear Agency (DNA). The 100-500 MJ unit may need more development due to lack of previous experiments. The 1-20 MJ unit is mobil, and similar to that of commercially available products. In most SMES applications, magnetic field has to be confined to insure the Air Force operation. It is suggested that further research should be conducted to transform existing technologies into practical SMES units suitable for the Air Force environment.

APPLICATIONS OF SMES IN AIR FORCE

Xingwu Wang

I. INTRODUCTION

Since the discovery of superconductivity in 1911, the idea of Superconducting Magnetic Energy Storage (SMES) had been proposed and studied many times.¹ Between 1983 and 1984, a 30 MJ SMES unit was installed and tested at the Tacoma Substation of the Bonneville Power Administration.² When high temperature superconductors were discovered several years ago, the concept of SMES was much more widely evaluated.^{3a-b} Up to date, small units (less than 10 MJ) are commercially available.^{4a-c}

As far as the SMES research is concerned, the driving force is mainly in the United States.^{4d} Within the Department of Energy (DOE), several National Laboratories have evaluated the energy storage capabilities of SMES.⁵ Electrical Power Research Institute (EPRI) has conducted independent research, and assisted government agencies on various SMES projects.⁶ Within the Department of Defense (DOD), different branches have looked into the SMES applications.⁷ Several years ago, the White Sands Missile Range examined the possibility of using SMES as a pulse energy source for the free electron laser, which was a part of the "star war" plan. The Navy studied the functions of SMES in ships; and the Army evaluated a conceptual design of SMES tanks. Currently, the Defense Nuclear Agency (DNA) is managing a feasibility study on an Engineering Test Model (ETM) of SMES (40 - 75 GJ range).⁸ The Army Corps of Engineers may play a major role in the design and engineering of the SMES ETM facility.⁹ From existing publications in the public domain,^{9a} we could not find any document dealing with the SMES applications in the Air Force, which is much more needed now as the SMES technology gradually matures.¹⁰ The objective of this paper is to provide preliminary information on the possible applications of SMES in the following six areas: (1) regular base, (2) bare base, (3) communications squadron, (4) air traffic control, (5) Sector Operational Control Center (SOCC) Radar station, and (6) energetic weapon. Other seven areas are considered, including space command and intelligence command.

II. ENERGY STORAGE AND POWER CONDITIONING

In an electrical system, a primary power supply normally provides electricity to the load. Occasionally, the primary supply may not be able to deliver the required quantity and quality of power due to various reasons. In order to maintain highly reliable power delivery, a secondary power supply or system should be activated, which may be another set of power lines, another generator, a power conditioner, or an Uninterruptible Power Supply (UPS). The emerging SMES technology also offers a cost effective approach to store energy and to improve power quality within a range of technical limits.

A. EXISTING UPS

There exist two kinds of UPS devices, rotational and stationary.^{10a} In the rotational UPS device, mechanical energy is stored in a flywheel. While in the stationary device, electrical energy is stored in a stack of batteries. Since 1988, the Air Force has installed battery UPS devices in more than one hundred air bases. The power of these devices varies from 1 KW to 1 MW,^{11a-b} and the battery discharge time varies from 10 to 20 minutes. The energy storage efficiency is approximately 70%. The lifetime of batteries and equipment is between five and fifteen years respectively. These devices are playing crucial roles in several areas including communication and Radar operation. The implementation of battery UPS technologies provides an opportunity to examine another complementary technology - SMES technology. In comparison with UPS batteries, SMES has following advantages: (a) high peak power output, (b) large energy storage capability, (c) long lifetime, (d) no chemical leakage, (e) no environmental problem.

B. SMES

In a SMES unit, a direct current (DC) flows in a superconductive coil, and energy stores in a magnetic form. The stored energy can be released to an external load via conventional AC/DC switching devices. The storage capability can be measured by the power and discharge duration. Based on existing studies, the SMES power varies from 460 KW to 1 GW, and the duration time varies from 0.25 seconds to 11 hours.^{4a-c, 12-13} (Other power levels and durations outside these ranges are also possible.) The response time to the load is usually less than 20 milliseconds. Energy dispatch efficiency is approximately 95%, and estimated lifetime is longer

than thirty years.¹³ The superior performance of the SMES is mainly due to the excellent properties of superconductors.

C. SUPERCONDUCTORS

A superconductor possesses two distinct features: zero resistance and Meissner effect at a temperature below its superconductive transition temperature, called critical temperature T_c .¹⁴ When a current is induced in a superconductive ring, a magnetic field is produced and the energy will be stored without loss. For a coil with inductance L , the stored energy, E , is given by:

$$E = L I^2 / 2 \quad (1)$$

where I is the magnitude of the current. As the current flows in magnetic field produced near the coil, Lorentz forces will be exerted onto the coil. In a solenoid, the coil will be pushed outward along radial directions, and inward along axial directions.

The magnetic field produced is not confined in a solenoid. To reduce magnetic field in solenoid configuration, even numbered solenoid coils can be assembled as shown in Fig. 1, where current flow in a coil is opposite to that of an adjacent coil. Besides solenoid, coil shape can be toroid, where magnetic field is confined.

If the current density in a superconductor exceeds a limit, called critical current density J_c , the superconductivity will be destroyed. The upper limit of J_c is approximately 10^6 A/cm² in zero magnetic field. If the magnetic field surrounding a superconductor exceeds a limit, the superconductivity will also be destroyed. There exists one critical magnetic field, H_c , in a type I superconductor such as a metal material.¹⁵ However, there exist two critical fields, H_{c1} and H_{c2} , in a type II superconductor such as an alloy or oxide material. In a type II superconductor, superconductivity is partially destroyed as soon as the magnetic field exceeds H_{c1} . Beyond H_{c2} , superconductivity is completely destroyed.

Therefore, several parameters should be considered when we select a superconductive material for SMES applications, i.e., T_c , J_c , H_c (H_{c1} , H_{c2}), and mechanical strengths. Commercially available materials for SMES coils are alloy superconductors including NbTi and Nb₃Sn.¹⁶ Rapid progress has been made on conductor technologies based on high T_c (oxide) superconductors. It is projected that we are at least 2-5 years away from making wires that are reliable enough for SMES coil applications.¹⁷ However, high T_c superconductors may be used as leads between low T_c superconductors and normal conductors in the very near future.^{17a1}

Since high T_c superconductors may reduce the cost and enhance the performance of SMES,^{3a-b}
⁵ a continuous support to the research is needed.^{17a2}

D. MAJOR COMPONENTS OF SMES

To illustrate operational principles of SMES, a block diagram is shown in Fig. 2. There are several components in a SMES unit, including superconductive coil assembly, DC-AC conversion assembly, cryogenic system, external system, and control system. The cryogenic system provides coolant to the superconductive coil assembly. The external system is normally an AC power grid. When the grid has excess electricity, an AC current is converted into a DC current via the DC-AC conversion system. The DC current is then supplied to the coil, and magnetic field is produced. When the current reaches a desired value, a switch closes the coil loop and the current will be circling along the loop. Thus, magnetic energy will be stored in the superconductive coil without loss. When the grid is overloaded, energy will be released from the superconductive coil assembly via a reversed process. Both storing and releasing processes are coordinated by the control system.

III. APPLICATIONS OF SMES

Traditionally, SMES is regarded as an energy storage unit, in which excessive electricity may be stored for late usage during a heavy load period. Alternatively, SMES can be used as a high power pulse source; in which electric energy provided by a conventional power supply is first accumulated over a period of time, and a high power pulse is then delivered. Recently, SMES of 500 KJ - 10 MJ is being used as a power conditioner, in which electric energy is stored and released to improve power quality. Traditional electrical devices do not require high quality power, while modern computers and electronics require high quality power since semiconductors are susceptible to changes in voltage and/or current.^{17b1} Poor power quality costs billions of dollars every year in the U.S. industry.^{17b2}

As far as the Air Force is concerned, power quantity and quality are important issues. The power quantity is normally regulated and maintained by the civil engineering personnel.^{17c} As of today, the power quality is mainly monitored by individual users. Since the power of a sensitive electronic equipment should have extremely high quality and reliability, a policy on power conditioning and continuation was issued four years ago.^{17d} Due to the nature of the Air

Force operation, problems in a power grid or distribution center are far more severe than that of an ordinary commercial site. In an Air Force installation, following problems need to be solved: (1) low quality in an original power source, (2) overload, and (3) interferences.^{17e} One of the solutions is to use SMES as an energy storage device or a power conditioner.

Applications of SMES include following items: (1) to reduce peak power load, (2) to prevent power outage, (3) to prevent voltage sag or swell, (4) to stabilize or alter frequency, (5) to prevent voltage surge, spike, or impulse, (6) to regulate voltage level, (7) to reduce harmonic distortion, (8) to eliminate electrical noise, and (9) to change the phase relation between voltage and current. Usually, SMES is connected to a power grid in a standby mode. It takes less than 20 mS to activate SMES, and the operation time of SMES varies from several millisecond to several hours depending on the nature of the usage. Some of these items will be discussed when we search for potential users in the Air Force.

A. REGULAR BASES

The electricity usage of a regular base can be obtained from the Defense Energy Information System (DEIS II) database. As an illustration, the energy consumption of 1991 is shown in Fig. 3, where the number of air bases is plotted as a function of the energy usage for 170 air bases. There are 63 bases with energy consumption less than 10 GW-Hr, in which 15 bases are in the range between 4 to 5 GW-Hr, see a blowup in Fig. 4. Referring to Fig. 3, there is a wide spread between 10 and 140 GW-Hr, in addition to a scattering in the higher energy region.

As an example, let us look at the Tyndall Air Force base, where the electricity usage in 1991 was 9.9356 GW-Hr. In Fig. 5, electrical power is expressed as a function of time for a typical summer day. The average power is approximately 15,500 KW, with peak load hours mainly in the day time and light load hours mainly in the evening. To reduce the peak load, a SMES of 20 MW-Hr may be connected to the power grid. The SMES receives energy during light load hours, and releases energy during the peak hours.

Since the energy storage requirement of 20 MW-Hr is very close to the ETM capacity (10 MW - 2 Hr), existing ETM studies should be applicable for our studies.⁸ For example, the diameter of the ETM coil is estimated to be 100 meters,¹³ which is shorter than the length of an airport runway. Thus, a regular base would be able to accommodate a SMES facility. If the

space above the ground is limited, the SMES could be installed beneath the runways.

Besides the peak load reduction, SMES can also be used in a regular base to prevent unexpected power outages.¹⁸ For example, during the storm seasons, bases in Florida, California, and other states experience power outages due to lightings.¹⁹ Most of these outages are momentary. If a SMES of 20 MJ is connected to the Tyndall grid, it will provide the emergency backup power to the base for 1-2 seconds.

In a regular base, the weight of a SMES unit is not limited. However, the magnetic field in the SMES has to be confined. The coil shape can be toroid, solenoid with magnetic shielding, or an assembly of even numbered solenoid coils.

B. BARE BASES

Guidelines for a bare base electrical power system are available in several publications.²⁰ A nominal electrical generator has a power of 750 KW, a voltage of 2400/4160 volts, and a frequency of 60 Hz, with three phases. The number of generators of each base depends on the load. The electrical wiring configuration also depends on the load. Usually, a normal bare base requires four generators in parallel, in addition to a standby generator. The Al Kharj base built for the "desert storm" has seventeen generators. The weight and size of any electrical system should be restricted to the limits of a C-130 aircraft, which is responsible for military deployment. The maximum payload of the plane is 44,000 pounds, the maximum per-axle load is 10,000 pounds.²¹ The length of the cargo space is 40 ft, and both width and height are 8 ft. In a bare base, the coolant for SMES, liquid helium, is not available and has to be produced by a close cycle refrigerator.

In a bare base power system, a backup generator is usually available to maintain reliable power. Before the backup takes over the full load, there may be a momentary power sag or outage. To avoid the power interruption, a SMES may be connected to the power grid in a standby mode. Its power level should be higher than 750 KW, and the discharge duration should be longer than 6 seconds.

Let us consider the existing commercial product as candidates for the application.⁴ First, the maximum energy storage capability of the product is approximately 2.5 MJ, which meets the minimum requirement. Second, the magnetic field surrounding SMES is large due to the design of an unshielded solenoidal coil, which is not acceptable in the bare base operation. To

overcome the problem, magnetic shielding may be used, even numbered solenoids may be used, or the toroidal coil may be used. However, the weight of SMES may be increased substantially. Based on a theoretical analysis, the coil assembly weight will increase 100% when a solenoidal design is changed into a toroidal design.²¹ The total weight of the existing commercial product is 25,000-30,000 pounds, which is below the load limit of the C-130 plane.^{4,23} If the coil weight is going to be increased, the weight of other equipments (such as air condition, control, and/or cryogenic assembly) should be decreased accordingly.

C. COMMUNICATIONS SQUADRON

Functions of a communication squadron include command, control, communication, and computer. Average power consumption of the squadron is approximately 300 KW. As more and more electronic/electrical equipments are being added, demand on electricity increases. If the main power supply fails, a standby or backup generator may be used to sustain the power supply. However, there will be a power-outage moment of approximately 6 seconds during switching.

Existing commercial SMES products (2.5 MJ) should be able to provide energy during momentary power outage. Ideal energy storage capacity for the squadron is 100 MJ. Since the energy requirement is relatively low, a hybrid system of SMES and battery UPS should be tested in the future. In a situation where battery disposal is prohibited, SMES may play an essential role. Magnetic field of SMES should be confined, and SMES facility could be installed in an underground compartment.

D. AIR TRAFFIC CONTROL

The average power of an air traffic control is 150 KW. If the control center is mobile, the major constraint is the weight. Existing commercial products should be able to provide energy when the primary supply is temporary out of service. If the center is stationary, the ideal energy storage capacity should be 20 - 60 MJ. The SMES could also be installed beneath the airport runway.

E. SECTOR OPERATIONAL CONTROL CENTER (SOCC) RADAR STATION

The main function of a SOCC Radar station is air surveillance in a specified region. Electricity supply to a critical load has to be continuous. The minimum power requirement is

approximately 540 KW. In order to provide this amount of the power for several seconds to several minutes, the energy capacity of a SMES should be in the range of 1-150 MJ. Magnetic field has to be confined near the SOCC stations.

F. ENERGETIC WEAPON

Several different kinds of modern weapons require substantial amounts of electrical power and/or energy. For example, a rail gun system has to discharge a large pulsed current within a short period of time. The power of a small rail gun is 1-2 MJ, with a peak current of 3 KA and a discharge time of 0.5 mS. While the power of a large rail gun is 500 MJ, with a peak current of 5 MA and a discharge time of 5 S. The existing power supply contains many batteries and/or capacitors, which is cumbersome and inefficient. In a large rail gun, capacitor banks need to be charged several times to reach peak power. Thus the gun firing speed and repetition rate are limited. If a SMES is adopted, the size of the power supply will be decreased, and the charging time will also be reduced. The required energy level of a SMES unit is in the range of 1-500 MJ. Magnetic field has to be confined in the rail gun applications.

G. OTHER AREAS

Since this study was carried out within eight weeks, we were unable to cover all important areas. Following examples are merely illustrative for future studies.

The first example is related to the large power applications in the space command, which is in charge of at least 18 bases and many sites. One of the responsibilities of the command is the satellite communication and/or operation. A conventional means for the satellite communication is a pulse Radar. The maximum distance between a satellite and a Radar is approximately 22,500 miles, and the minimum power required is approximately 6-20 MW. It is difficult to reach the peak power requirement with existing battery and/or capacitor banks. When a pulse Radar is in operation, the power grid may be overloaded and other electronics may be affected. A study for SMES applications could be carried out in Cheyenne Mt. AFB, Falcon AFB, or Onizuka AFB.

The second example is related to the small power applications. In every air base, there is a command post which is a focal point of the operation. The average power of the post is approximately 100 KW. A case study should be carried out to examine the compatibility of

SMES and battery UPS.

The third example is related to frequency stabilization and alteration where SMES is used as a direct current source. The frequency stabilization is a part of power conditioning processes. While the frequency alteration is required when an Air Force unit is operating in Foreign countries where frequency is 50 Hz with voltages of 230/380 or 240/416 volts.

The fourth example is related to a SMES unit for an aircraft. Since 1960's, several inductive energy storage techniques have been tested for pulse power applications.^{9a, 24a} In a modern surveillance airplane, more electronic instrumentations are being added, and more generators and batteries are being added accordingly. As a result, the power quality deteriorates, and ultimate surveillance goal may not be reached. To improve the power quality, a SMES may be used as a power conditioner to generate high power pulses. The maximum SMES energy should be 1 MJ or less. The weight and size of SMES should be less than that of a battery unit. An earlier study sponsored by NASA could be used as a reference, where SMES was considered for an application in a space shuttle.^{24b1}

The fifth example is related to the overseas installations such as Pacific Air Forces and United States Air Forces in Europe. In comparison to the domestic bases, power quality is usually poorer, and power quantity is usually larger. For example, the power of a consolidated or centralized computer may be 1 MW or larger. At this power level, SMES would be much more suitable than the battery UPS. Furthermore, if battery disposal is prohibited in a Foreign country, the shipping cost would be very expensive. A case study may be carried out in Ramstein AFB or Yokota AFB.

The sixth example is related to the intelligence command, where information gathering and distribution are heavily dependent on the computers and electronic instrumentations. The power of a critical load is estimated to be 500 KW - 2 MW, or higher. A case study should be carried out in Kelly AFB.

The seventh example is related to the prevention of high-altitude electromagnetic pulse (HEMP). These pulses are produced by radiations of nuclear device (such as gamma rays) in the atmosphere at altitudes above 20 km. The peak value of a pulse is approximately 50 KV per meter. When the HEMP strikes an electrical network or system, the maximum current can be as high as 8,000 A, the pulse duration can be as long as 100 S. Usually, a HEMP filter is connected to the electrical network to dissipate the pulse energy.^{24b2} If SMES is used, the HEMP

energy can be stored for late usage.

IV. DISCUSSIONS AND RECOMMENDATIONS

After examining several areas, we believe that the SMES technology is applicable in the Air Force. The required energy will be in three ranges: 20 MW-Hr (72 GJ), 100-500 MJ, and 1-20 MJ. The required discharge time will be from several mS to several hours.

Our recommendations to the Air Force include both short and long term plans. Since this study is mainly concentrated on the technology availability, the cost of a SMES unit is not the deterministic factor in the recommendations.^{24c} (The cost would be decreased as the industry matures.)

A. SHORT TERM PLAN

(1) The Air Force should continuously monitor the progress of SMES applications in other government branches and industries. In particular, the Air Force should follow the ETM studies (10 MW - 2 Hr), which would be applicable in a regular base. There are several sites proposed for the ETM installation: Monahans, Texas; Baraboo, Wisconsin; Orogrande, New Mexico; White Sands Missile Range, New Mexico; and Richland, Washington. Some of these sites are close to the existing Air Force bases including Cannon AFB, Holloman AFB, Kirtland AFB, and Reese AFB. After the demonstration experiment is completed, the ETM unit will be useful for the nearby Air Force base.

(2) The Air Force should look into the design of a small SMES in the 10-20 MJ range, which would be movable by a C-130 plane. The objectives are: (a) to reduce the weight and size; (b) to reduce the magnetic field leakage; (c) to enhance the structure integrity.

(3) In the SBIR programs, the Air Force should support a small scale experiment which must be unique to the Air Force applications. If results in phase I are fruitful,¹⁰ the Air Force should support a phase II study, in which the experiment could be carried out. An ideal base for the study should be located in a state where lightnings occur often, such as California or Florida. To simulate a realistic situation in the air force, the SMES unit should be installed in a training center, a research laboratory, and/or a mainframe computer room. As an example, the power of a mainframe computer at Tyndall Air Force base is 40 KW, with three phases and 208 volts. During the experimentation, engineers from aforementioned units should be invited to inspect

SMES operations. If phase II project is successful, phase III project may be conducted in one of the aforementioned units, where engineers are familiar with the SMES.

B. LONG TERM PLAN

The effort on SMES should benefit to the ultimate goals of Air Force operations. If the SMES technology is implemented in the near future, following problems should be considered during engineering design stages.

(1) Magnetic field confinement of a SMES facility should be considered to secure a military operation even though there is no military standard available. In addition, human health may be affected by strong magnetic field as reported in several publications. Without a safety guideline from health organizations, the reference of the exposure level can be earth magnetic field which is 0.5 - 1 Gauss. We suggest that the field should be lower than 10 Gauss at a distance of 1 meter away from the enclosure of the SMES coil. This limit is achievable with existing technologies.²⁵

(2) Since movable SMES units will be carried by cargo planes, we suggest that existing military standards for electrical generators should be used as references for the design of an air transportable SMES unit.²⁰⁻²¹ During initial experiments with ground transportation, existing military specifications for truck and trailer transport of mobile electric generators should be used as references.²⁴ A SMES standard should be proposed before the permanent implementation takes place.

C. COMMUNICATIONS WITH OTHER AGENCIES IN THE FIELD

Because SMES is a new technology, the Air Force should exchange information with other agencies in the field. The R&D manager(s) should attend following SMES meetings. A meeting organized by DOE for DNA will be held in September 1992 (Milwaukee, Wisconsin). Another meeting, organized by EPRI, will be held in October (Palo Alto, California). It is desirable to set up a SMES government interest group meeting due to sufficient demand in militaries, national laboratories, energy administration offices in both federal and state governments. Either DOE or DNA could take a lead to organize the meeting.

The Air Force should also keep in touch with following companies which are associated with SMES designs and fabrications: General Electric; Westinghouse; Intermagnetics General

Corporation; General Atomic; Superconductivity, Inc; General Dynamics; Bechtel National, Inc.; Advanced Cryomagnetics, Inc.; EBASCO Services Incorporated; American Superconductor Corp., W. J. Schafer Associates; etc.

Connections between the Civil Engineering Laboratory at Tyndall Air Force base and the Aero-Propulsion Laboratory at Wright-Patterson Air Force base are beneficial. Even though the responsibility of each laboratory is different, both laboratories are pursuing superconductor applications in the Air Force. Two laboratories may consider collaborative research on superconductivity in the future.

Regarding to the electricity management system, existing civil engineering support is suitable for the power delivery as far as the quantity is concerned. However, power quality control may be a potential problem, and may require a parallel study in connection to SMES applications.

ACKNOWLEDGEMENT

The author would like to thank the following personnel at Tyndall Air Force Base for their help during the study: Ed Alexander, Tom Hardy, Capt Lori Young, Resa Salavani, Tsgt Gaston Glassgow, Larry Strother, Andrew Poulis, Fred Beason, Lt Juoah Bradley, and David Conkling. In addition, Tom Hardy provided valuable suggestions to revise the manuscript.

REFERENCES

1. R. W. Boom, and H. A. Peterson, "Superconductive Energy Storage for Power Systems", IEEE Transactions on Magnetics, Mag-8, 701 (1972).
2. J. D. Rogers, "Superconducting Magnetic Energy Storage (SMES) Program", Los Alamos Progress Report, LA-10399-PR.
- 3a. S. M. Schoenung, et al, "Conceptual Design Study of Superconducting Magnetic Energy Storage using High Temperature Superconductors", Final report on Contract DE-AC02-90CE34019, February, 1992.
- 3b. R. J. Loyd, A. M. Bulc, and D. Majumdar, "Superconducting Magnetic Energy Storage: Technical Considerations and Relative Capital Cost Using High-Temperature Superconductors", EPRI TR-100557, April 1992.
- 4a. C. C. DeWinkel, and P. F. Koeppel, "Ensure Continuous Power to Critical Industrial Processes with the New Superconducting Magnetic Energy Storage System", presented at the Fourteenth National Industrial Energy Technology Conference, Houston, Texas, April 1992.
- 4b. A. K. Kalafala, L. Salasoo, F. G. Turnbull, and R. A. Ackermann, "A Small Superconducting Magnetic Energy Storage System for Utility Applications", GE Research & Development Center, June 1992.

- 4c. Private Communications with J. Porcell.
- 4d. Japan has two major SMES programs; Canada has at least two; Germany has at least one; and Switzerland has one. In addition, China has at least two programs related to SMES.
5. D. M. Rote, J. S. Herring, and T. P. Sheahen, "Transportation", in "Energy Applications of High-Temperature Superconductivity, Volume 2", ed. S. Dale, et al., (Electric Power Research Institute, Inc., Palo Alto, CA, 1990), p 5-1.
6. "Superconducting Magnetic Energy Storage Pilot Plant", EPRI Technical Brief, RP2988, TB.GS.90.2.90.
7. Private Communications with D. Jacobs, S. Wolf, J. Moya, M. Supercynski, and G. Guazzoni.
8. See, for example, "Superconducting Magnetic Energy Storage Project", EBASCO Services, Inc., Contract DNA001-92-C-0064.
9. Superconductor Week, July 6, 1992, p 4.
- 9a. We did find several abstract of internal reports on inductive energy storage systems from Wright-Patterson AFB during 1964 - 1976. Per a private conversation with C. Oberly, we learned that his group is mainly working on conductors which will be useful for energy storage.
10. In June 1992, the Air Force awarded SBIR grants to two companies who will conduct research on the critical needs of SMES.
- 10a. "Status of Static UPS Applications in the United States", EPRI CU-6498, (Electric Power Research Institute, Inc., Palo Alto, CA 1989).
- 11a. "Power Conditioning and Continuation Interfacing Equipment Handbook".
- 11b. When the battery power exceeds 1 MW, the total weight of UPS is beyond the load capacity of a C-130 plane.
12. The maximum power delivery is extracted from a conceptual design; see, for example "Conceptual Design and Cost of a Superconducting Magnetic Energy Storage Plant" by R. J. Loyd, et al., EPRI EM-3457, 1984.
13. R. B. Schainker, "Superconducting Magnetic Energy Storage (SMES): Perspectives, Prospects and Programs (an EPRI position Paper)", EPRI Gray Report GS-RBS-12/90.
14. See, for example, "Foundations of Applied Superconductivity" by T. P. Orlando and K. A. Delin, (Addison-Wesley Publishing Co., Reading, Massachusetts, 1991).
15. M. Tinkham, "Introduction to Superconductivity", (Krieger Publishing Co., Malabar, Florida, 1980).
16. These alloy superconductors are type II superconductors. See, for example, "Principles of Superconductive Devices and Circuits" by T. Van Duzer and C. W. Turner, (Elsevier, New York, 1981).
17. Private communications with D. Shaw, and E. Schempp.
- 17a1. Superconductor Week, July 27, 1992, p 2.
- 17a2. Private communications with R. Snyder.
- 17b1. "A Guideline on electrical power for ADP (automatic data processing) installations", FIPS PUB 94, U.S. Department of Commerce, September 1983.
- 17b2. "Problems with Power Quality", EPRI Journal, July/August, 1991; R. D. Hof, "The 'Dirty Power' Clogging Industry's Pipe Line". p 82, Business Week, April 8, 1991.
- 17c. Electricity in a lot of installations is purchased from local power companies. Other installations use self-generated electricity.
- 17d. "Policy for power conditioning and continuation interfacing equipment (PCCIE), Department of the Air Force, December 1988.

- 17e. When a battery UPS unit is connected to a load, the power quality of the load is improved. However, the overall quality of the grid decreases, and other electronics (computers) unprotected by UPS may not be functioning properly.
18. A recent test on a small SMES unit was carried out at IBM in New York State, where the SMES provided emergency backup to a VLSI manufacture facility during a power outage event caused by lightings. See "Superconductor Week", June 15, 1992, p 3.
19. As the report was being drafted, several lightings struck the Tyndall Air Force base and caused temporary power outages.
20. "Bare Base Conceptual Planning Guide", draft, third edition, AFESC, May 1985; HQ AFESC, SON 001-86, "Air Force Statement of Operational Need for Bare Base/ Backup Power Systems", May 1986; "Bare Base Power Supply Systems", AFESC Report ESL-TR-87-25, June 1988.
21. MIL-A-8421F, "Air Transportability Requirements, General Specifications For", 25 October 1974.
22. J. R. Hull, S. M. Schoenung, D. N. Palmer, and M. K. Davis, "Design and Fabrication Issues for Small-Scale SMES", Advances in Cryogenic Engineering, Vol. 37, Part A, ed. R. W. Fast, (Plenum, New York, 1992), p 369.
23. Private communication with C. C. DeWinkel.
24. MIL-STD-633E, "Mobile Electric Power Engine Generator Standard Family", 22 February 1980.
- 24a. We only have an abstract as mentioned in Ref. 9a. Final report No. AFAPL-TR-75-60, "Design, Construction, and Testing of a Pulsed High Energy Inductive Superconductive Energy Storage System". Work was performed by Magnetic Corp of America.
- 24b1. Private communications with R. Boom; Y. M. Eyssa, R. W. Boom, G. E. McIntosh, and Q. Li, "A 100 KWh Energy Storage Coil for Space Application", IEEE Transactions on Magnetics, Vol. Mag-19, 1081(1983); Y. M. Eyssa, R. W. Boom, X. Huang, M. A. Hilal, M. K. Abdelsalam, L. O. El-Marazki, and M. J. Superczynski, "Superconductive Magnetic Energy Storage (SMES) for Space Applications", Proceedings of European Space Power Conference, September 1991, Florence, Italy.
- 24b2. G. Champman, and A. Chodorow, H. Pohle, and R. Berglund, "USAF Handbook for the Design and Construction of HEMP/TEMPEST Shielded Facilities, Mission Research Corp., Albuquerque, NM, December 1986, Revised July 1988.
- 24c. In 1991, the SMES funding levels of 13 laboratories varies from nineteen thousand dollars to five million dollars, with a total amount over 6.6 million dollars. See, for example, L. Dresner, "Survey of Domestic Research on Superconducting Magnetic Energy Storage", ORNL HTSPC-1, 1991.
25. Private communications with K. Kalafala.

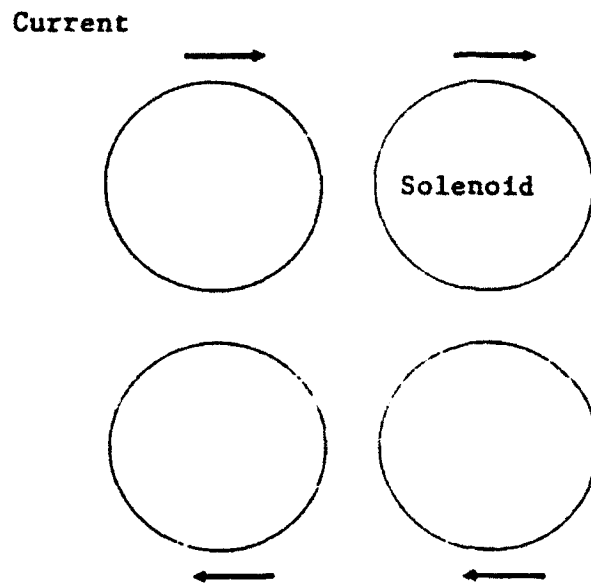


Fig. 1. Even numbered solenoids to reduce stray magnetic field

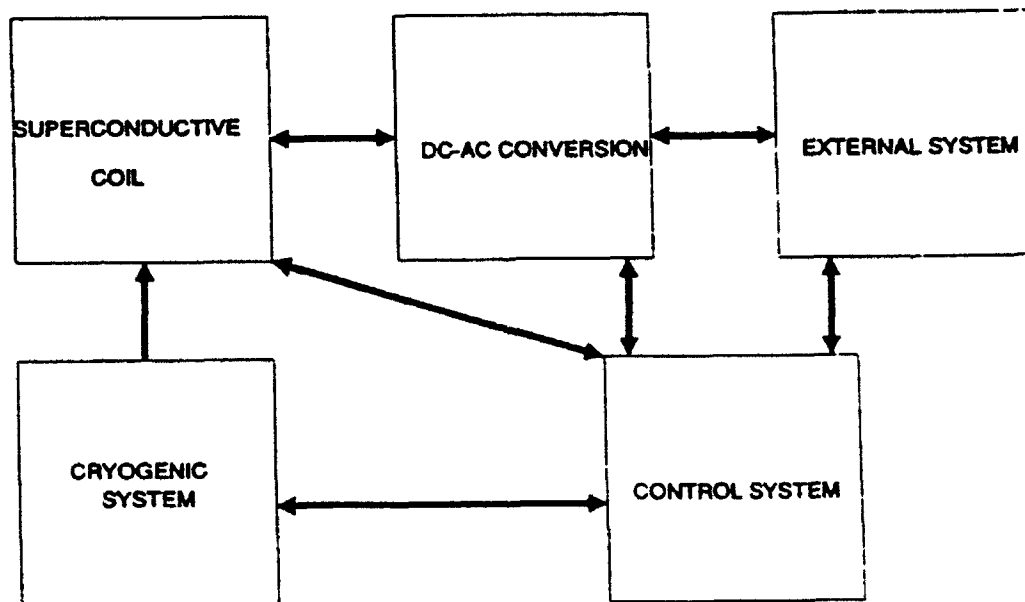


Fig. 2. Major components of SMES

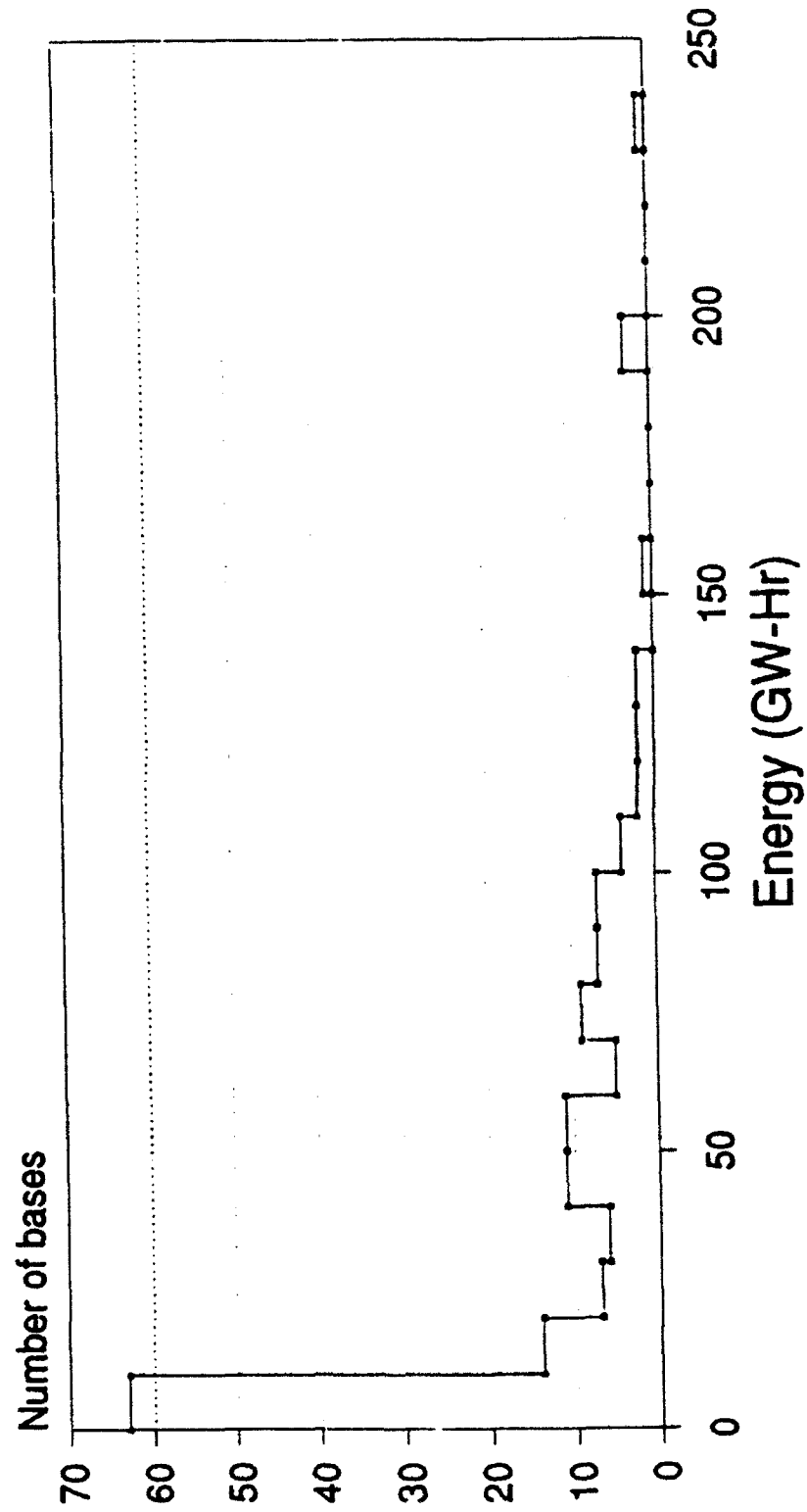


Fig. 3. Annual electric energy usage (1991)

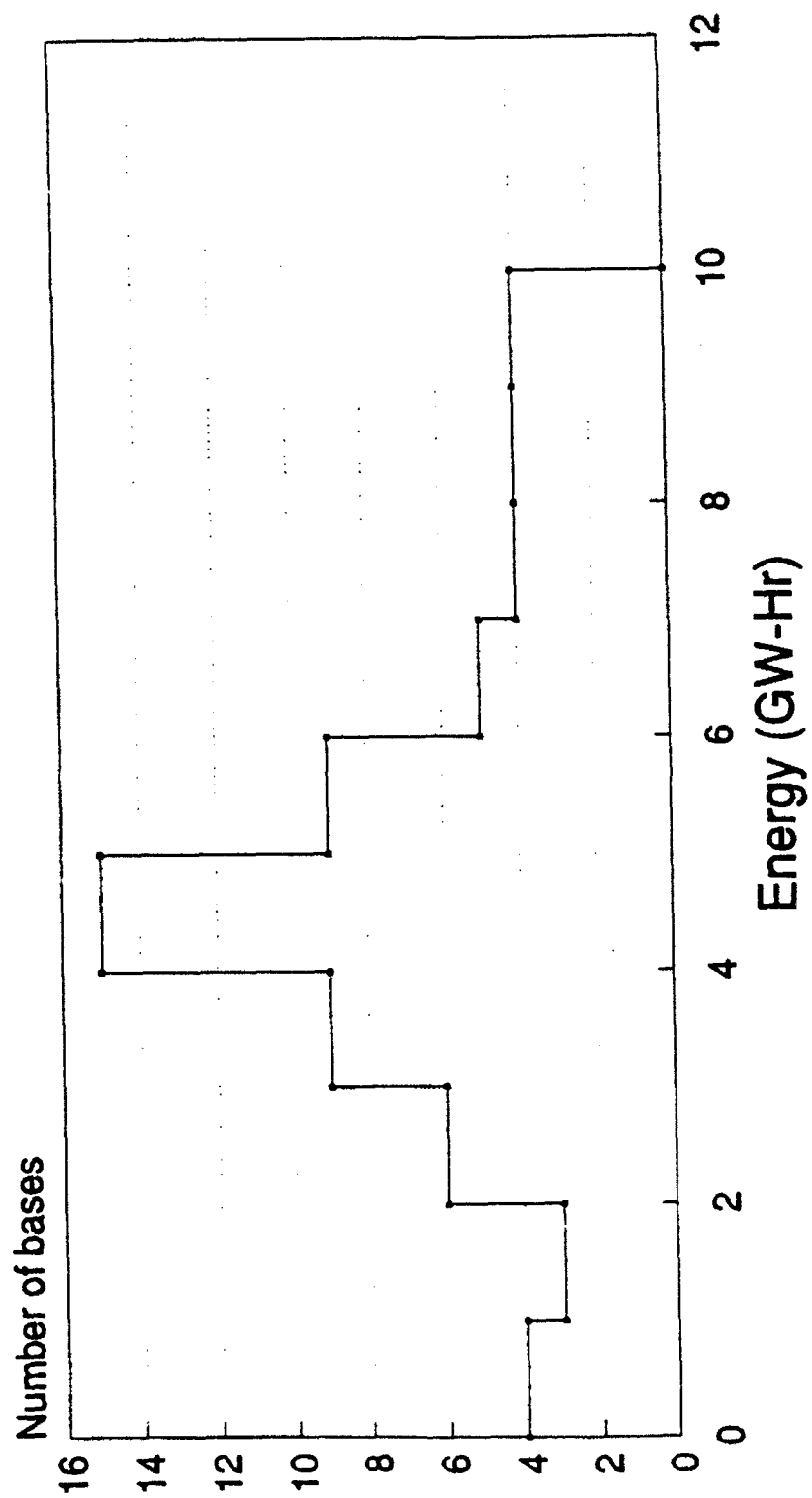


Fig. 4. Annual electric energy usage (1991)

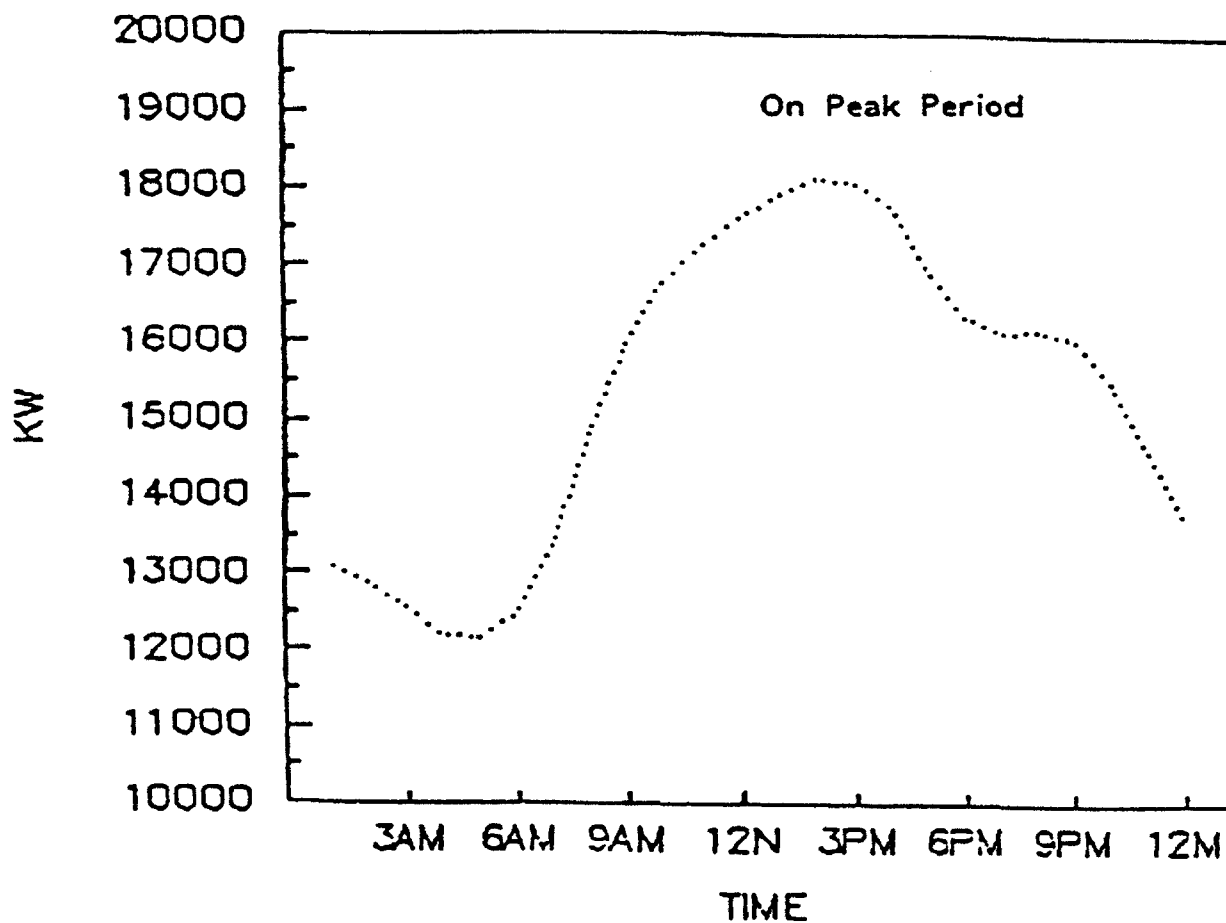


Fig. 5. Electricity load for a typical summer day at Tyndall AFB

FRANK J. SEILER RESEARCH LABORATORY

MULIVARIABLE TRANSFER FUNCTIONS AND OPTIMAL PASSIVE DAMPING FOR A SLEWING PIEZOELECTRIC LAMINATE BEAM

Thomas E. Alberts
Associate Professor

and

Travis DuBois
Graduate Student

Department of Mechanical Engineering and Mechanics
Old Dominion University
Norfolk, Virginia 23529-0247

Final Report for:
Summer Research Program
Frank J. Seiler Research Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

July 31, 1992

MULTIVARIABLE TRANSFER FUNCTIONS AND OPTIMAL PASSIVE DAMPING FOR A SLEWING PIEZOELECTRIC LAMINATE BEAM ¹

Thomas E. Alberts & Travis V. DuBois
Department of Mechanical Engineering and Mechanics
Old Dominion University
Norfolk, Virginia 23529-0247, USA
Phone: (804) 683-3736
Email: taiberts@mem.odu.edu

Abstract

This report presents the development and experimental verification of a distributed parameter model for a slewing beam system with piezoelectric actuators and sensors. The beam is pinned at the proximal end, an endpoint motion sensor is attached at the distal end, and patches of thin piezoelectric laminates attached to its surface. The differential equation of motion for this system is transformed to Laplace domain transfer functions after application of the appropriate boundary conditions. Transfer functions relating the various actuator/sensor pairs are developed. The transfer functions are rationalized using a Maclaurin series expansion so that there is no need to assume mode shapes. Experimental results, which verify the model, are presented using a beam experiment at the US Air Force Academy, Frank J. Seiler Research Laboratory. The existing clamped beam experiment was modified through the addition of a hinged joint and appropriate instrumentation to carry out this work.

The transfer functions are eventually to be used to develop and experimentally validate a simultaneously optimal active and passive damping design for the experimental system. A preliminary damping design is discussed and initial experimental results presented.

¹ This work performed in collaboration with Dr. H.R. Pota of the Australian Defence Force Academy.

MULTIVARIABLE TRANSFER FUNCTIONS AND OPTIMAL PASSIVE DAMPING FOR A SLEWING PIEZOELECTRIC LAMINATE BEAM

Thomas E. Alberts & Travis V. DuBois

1 Introduction

Piezoelectric laminates are under investigation as actuators and sensors for vibration control in flexible structures. Piezoelectric materials in current use include poly-vinylidene fluoride (PVDF), a semicrystalline polymer film, and lead zirconate titanate (PZT), a piezoelectric ceramic material. These materials strain when exposed to a voltage and, conversely, produce a voltage when strained. This property exists because of a permanent dipole nature of the materials, induced by exposing it to a strong electric field during manufacture. Detailed discussion of the unique electro-mechanical nature of these materials is available in [2]. Piezoelectric layers bonded to the surface of, or manufactured into flexible structure members can act as either control actuators or sensors. A unique characteristic of PVDF and PZT actuators and sensors is that they are spatially distributed over the surface which is being controlled. This is in contrast to discrete actuators and sensors more customarily employed in the control of structures. Perhaps due to this spatially distributed nature, piezoelectric laminate structures are seldom analyzed using classical transfer function models. However, because the spatial distribution of the control influence or sensing capability is fixed once the laminates are applied, each actuator is associated with a discrete input variable and each sensor a discrete output variable (both are voltages) thus transfer functions can be derived.

One of the first published accounts of research using piezoelectric laminates for structural control was produced by Bailey and Hubbard [4]. They developed a dynamic model for a cantilevered beam with a PVDF film actuator, by incorporating the piezoelectric electro-mechanical relationships into the Bernoulli-Euler beam equation. Finite element analysis was applied by Baz and Poh [5] and the assumed modes method by Garcia, Inman, and Dosch [7], to model piezoelectric laminate beams. Pan, Hansen, and Snyder [6] analyze a pinned-pinned flexible beam with one piezoelectric element covering only a part of the beam surface. The system is modelled using two separate Bernoulli-Euler equations, (for x such that $0 < x < x_1$ and $x_2 < x < l$, cf. Figure 2) with no forcing input, and the effect of the piezoelectric element is incorporated as eight separate boundary conditions. In a previous work [3] the first author presented transfer function models for a cantilevered beam with PVDF film actuator/sensor pairs covering its entire surface. This report firstly extends that work to a slewing beam with actuator/sensor pairs which cover the surface partially and secondly presents experimental results demonstrating the validity of this modelling technique.

This report is organized as follows. In the next section the electro-mechanical relationships required for modelling piezoelectric laminates are reviewed. In section 3 these relationships are incorporated into a dy-

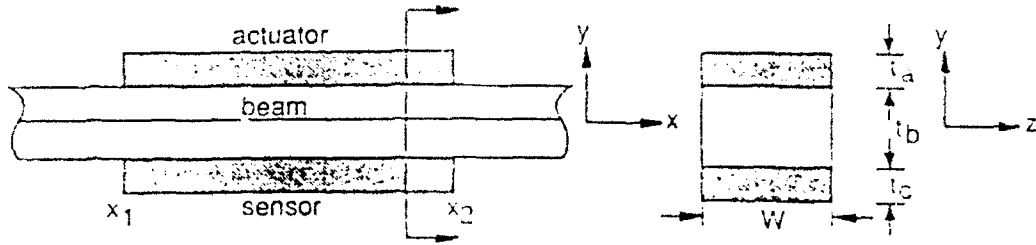


Figure 1: Schematic of Piezoelectric Laminate

dynamic model for the piezoelectric laminate beam. This model is used to derive exact transfer functions, given as transcendental functions of a modified frequency parameter β , for a general flexible slewing piezoelectric laminate beam. Section 4 gives a finite truncation of the exact transfer functions, derived in the previous section, as rational functions of the Laplace operator s . In section 5, the 2 m long experimental piezoelectric laminate beam at the US Air Force Academy, Frank J. Seiler Research Laboratory is described. Section 6 presents the experimental results which verify the exact transfer functions derived in this report. Section 7 presents some preliminary passive damping results.

2 Piezoelectric Laminate Electromechanical Relationships

Figure 1 illustrates a typical section of the beam/piezoelectric layer laminate. The subscripts a , b , and c correspond to the top piezoelectric layer, the base structure and the lower piezoelectric layer respectively.

In the present development, the a layer will serve as actuator and the c layer as sensor. A voltage V_a applied across the actuating layer a induces a longitudinal stress σ_f given by:

$$\sigma_f(x, t) = E_a \frac{d_{31}}{t_a} V_a(x, t) \quad (1)$$

where d_{31} is the electric charge constant of the film (m/v), and E_a is the Young's modulus of the film (N/m^2). This stress in turn generates a bending moment M_{f_a} about the composite system's neutral axis, given by [5]:

$$M_{f_a}(x, t) = \int_{-\frac{1}{2}}^{\frac{1}{2}+t_a} \sigma_f w y dy = \frac{1}{2} E_a d_{31} b (t_a + t_b) V_a(x, t) = C_a V_a(x, t) \quad (2)$$

where $C_a = \frac{1}{2} E_a d_{31} w (t_a + t_b)$ is a geometric constant, the form of which is determined by the geometry of the composite beam-actuator-sensor system. The constant changes if, for example, both the a and c layers are used as actuators with opposite polarity, then $C_a = E_a d_{31} w (t_a + t_b)$. It is also worth noting that the expression for C_a is significantly simplified by the fact that the laminated beam's neutral axis coincides with its center due to the symmetry of the added piezoelectric elements.

Next an expression is obtained for the voltage across the sensing layer. Note that the strain ϵ_c is related to the curvature of the beam in the following way:

$$\epsilon_c(x, t) = - \left(\frac{t_b}{2} + t_a \right) \frac{\partial^2 y}{\partial x^2} \quad (3)$$

Due to the piezoelectric effect this strain gives rise to a charge distribution per unit area, $q(x, t)$, which is given by:

$$q(x, t) = \left(\frac{k_{31}^2}{g_{31}} \right) \epsilon_c(x, t) \quad (4)$$

where k_{31} is the piezoelectric electromagnetic coupling constant and g_{31} is the piezoelectric stress constant. The total charge developed on the sensing layer is obtained by integrating $q(x, t)$ over the entire surface area of the piezo element,

$$Q(t) = \int_{x_1}^{x_2} w q(x, t) dx = -w \left(\frac{t_b}{2} + t_a \right) \frac{k_{31}^2}{g_{31}} \frac{\partial y(x, t)}{\partial x} \Big|_{x_1}^{x_2} \quad (5)$$

The piezoelectric material is similar to an electric capacitor and the voltage across the two layers is given by the following formula:

$$V_s(t) = \frac{Q(t)}{C w (x_2 - x_1)} = C_s \frac{\partial y(x, t)}{\partial x} \Big|_{x_1}^{x_2} \quad (6)$$

where C_s is a constant incorporating various structural and piezoelectric constants in (5) and (6), C is the capacitance per unit area, and $w(x_2 - x_1)$ is the piezoelectric element's surface area. In the next section a dynamic modelling method for piezoelectric laminate beams is presented using the relationships given in this section.

3 Dynamic Modelling

Figure 2 illustrates the configuration of the system under consideration. It is regarded as a generic model representing slewing flexible beams with piezoelectric actuators and sensors. The beam is mounted on rotary base, actuated by a torque motor, and rotating in the horizontal plane. The piezoelectric patches are attached to both sides of the beam. In this development, one side acts as an actuator and the other acts as a sensor. Thus the system inputs consist of voltage V_a applied to the actuating layer and the torque T applied by the torque motor. Outputs include the voltage V_s sensed by the sensing layer, motor hub angle θ and beam tip position Y_{tip} .

Using the usual assumptions for technical beam theory, the differential equation of motion for the composite beam-film system can be expressed as a Bernoulli-Euler beam equation with an additional term due to the actuating layer [4]. The equation is given as

$$\frac{\partial^2}{\partial x^2} \left[EI \frac{\partial^2 y(x, t)}{\partial x^2} - C_a V_a(x, t) \right] + \rho A \frac{\partial^2 y(x, t)}{\partial t^2} = 0 \quad (7)$$

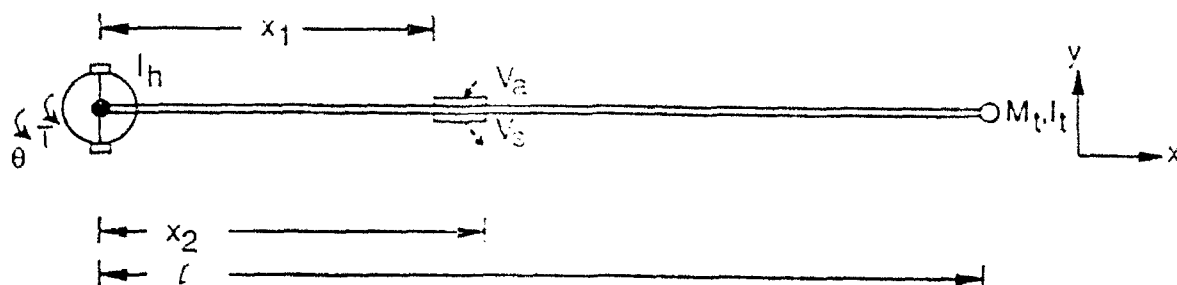


Figure 2: Schematic of Slewing Beam System

where E and I represent respectively the effective Young's modulus and area moment of inertia for the composite system. Notice that if the piezoelectric laminates do not cover the entire beam surface, then both EI and ρA are functions of x . In the laminated portions of the beam, the product EI is given by $EI = E_a I_a + E_b I_b + E_c I_c$, and the product of the effective mass density ρ with the composite system's cross sectional area A is $\rho A = \rho_a A_a + \rho_b A_b + \rho_c A_c$. Since the piezoelectric layers are often thin by comparison to the base structure, the present analysis is simplified by assuming that EI and ρA are uniform over the length of the beam. The boundary conditions are:

$$y(0, t) = 0 \quad (8)$$

$$EI \frac{\partial^2 y(0, t)}{\partial x^2} - I_h \frac{\partial^3 y(0, t)}{\partial t^2 \partial x} + T(t) = 0 \quad (9)$$

$$EI \frac{\partial^2 y(l, t)}{\partial x^2} + I_t \frac{\partial^3 y(l, t)}{\partial t^2 \partial x} = 0 \quad (10)$$

$$EI \frac{\partial^3 y(l, t)}{\partial x^3} - M_t \frac{\partial^2 y(l, t)}{\partial t^2} = 0 \quad (11)$$

To allow conversion to transfer functions, the differential equation (7) and boundary conditions (8)–(11) are Laplace transformed. Letting $\beta^4 = -\rho A s^2 / EI$, (7) becomes

$$Y''''(x, s) - \beta^4 Y(x, s) = \frac{C_a V_a''(x, s)}{EI} \quad (12)$$

with boundary conditions:

$$Y(0, s) = 0 \quad (13)$$

$$EI Y''(0, s) - I_h s^2 Y'(0, s) + T(s) = 0 \quad (14)$$

$$EI Y''(l, s) + I_t s^2 Y'(l, s) = 0 \quad (15)$$

$$EI Y'''(l, s) - M_t s^2 Y(l, s) = 0 \quad (16)$$

where the primes indicate spatial derivatives. Together, equations (12)–(16) constitute a linear ordinary differential equation in x with mixed boundary conditions — two at $x = 0$ and two at $x = l$. The differential

equation (12) is a Laplace transformed version of a classical Bernoulli-Euler beam equation with the second spatial derivative of $V_a(x, s)$ as a forcing input. Notice that when the spatial distribution of the actuating layer is uniform with respect to x , then the voltage term has a non-zero contribution only at the boundaries of the actuating layers. The actuating voltage V_a is constant in the interval $x_1 < x < x_2$, but undergoes a step change at each of the boundaries of this interval. Thus, the second spatial derivative of the actuating voltage function yields,

$$V_a''(x, s) = [\delta'(x - x_1) - \delta'(x - x_2)]V_a(s) \quad (17)$$

where $\delta'(\cdot)$, the spatial derivative of the dirac delta, represents the unit dipole function. It can be seen that, in effect, the actuating layer applies moments to the beam at its endpoints x_1 and x_2 . For a system with N piezoelectric actuator/sensor pairs a state-space type representation can be formed:

$$\begin{bmatrix} Y'(x, s) \\ Y''(x, s) \\ Y'''(x, s) \\ Y''''(x, s) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \beta^4 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y(x, s) \\ Y'(x, s) \\ Y''(x, s) \\ Y'''(x, s) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \frac{C_a V_a(s)}{EI} \sum_{i=1}^{2N} \delta'(x - x_i) (-1)^{i+1} \quad (18)$$

Defining

$$Z(x, s) \triangleq [Y(x, s) \ Y'(x, s) \ Y''(x, s) \ Y'''(x, s)]^T$$

and

$$U(x, s) \triangleq \frac{C_a V_a(s)}{EI} \sum_{i=1}^{2N} \delta'(x - x_i) (-1)^{i+1}$$

equation (18) can be written in compact form with appropriate definitions of A and B as follows:

$$Z'(x, s) = AZ(x, s) + BU(x, s) \quad (19)$$

Then using the fact [10] that,

$$\int_{-\infty}^{\infty} \delta^n(x) \phi(x) dx = (-1)^n \phi^n(0)$$

equation (19) can be expressed for $N = 1$, (i.e. only one piezoelectric actuator/sensor pair) as:

$$Z(l, s) = e^{Al} Z(0, s) + [Ae^{A(l-x_1)} B - Ae^{A(l-x_2)} B] \frac{C_a V_a(s)}{EI} \quad (20)$$

Now combining (20) with the boundary conditions (13)–(16) yields 8 linear algebraic equations in 8 unknowns, which can be solved simultaneously for the elements of $Z(0, s)$ and $Z(l, s)$. Once the end conditions $Z(0, s)$ and $Z(l, s)$ are known a general expression for $Z(x, s)$ can be written as follows:

$$Z(x, s) = \begin{cases} e^{Ax} Z(0, s), & 0 \leq x \leq x_1 \\ e^{Ax} Z(0, s) + Ae^{A(x-x_1)} B \frac{C_a V_a(s)}{EI}, & x_1 \leq x \leq x_2 \\ e^{Ax} Z(0, s) + A[e^{A(x-x_1)} - e^{A(x-x_2)}] B \frac{C_a V_a(s)}{EI}, & x_2 \leq x \leq l \end{cases} \quad (21)$$

The matrix exponential function e^{Ax} corresponding to (18) can be expressed as

$$e^{Ax} = \begin{bmatrix} f'''(x) & f''(x) & f'(x) & f(x) \\ \beta^4 f(x) & f'''(x) & f''(x) & f'(x) \\ \beta^4 f'(x) & \beta^4 f(x) & f'''(x) & f''(x) \\ \beta^4 f''(x) & \beta^4 f'(x) & \beta^4 f(x) & f'''(x) \end{bmatrix} \quad (22)$$

where, $f(x) = \frac{1}{2\beta^3} [\sinh(\beta x) - \sin(\beta x)]$.

The solution expressed above by (21) has been derived for a single piezoelectric element, however, because of the system's linearity, superposition allows the same analysis to be applied to systems with any number of piezoelectric elements. In the analysis which follows, transcendental expressions are developed for the transfer functions of interest. To reduce the complexity of the expressions, the tip mass M_t and inertia I_t are set to zero for the remainder of the report.

3.1 Transfer Functions

The system under consideration has two discrete inputs, V_a and T , and three outputs, Y_{tip} , θ and V_s . Thus, the transfer matrix can be expressed:

$$\begin{bmatrix} \theta(s) \\ V_s(s) \\ Y_{tip}(s) \end{bmatrix} = \begin{bmatrix} G_{\theta, V_a}(s) & G_{\theta, T}(s) \\ G_{V_s, V_a}(s) & G_{V_s, T}(s) \\ G_{Y, V_a}(s) & G_{Y, T}(s) \end{bmatrix} \begin{bmatrix} V_a(s) \\ T(s) \end{bmatrix} \quad (23)$$

The subscripts p and q of each element $G_{p,q}(s) = N_{p,q}(s)/D(s)$ identify the associated output and input variable respectively. Each of the six transfer functions for this system have the common denominator $D(s)$ given below.

$$D(s) = 2\beta^2 EI (\rho A (\cos(\beta l) \sinh(\beta l) - \cosh(\beta l) \sin(\beta l)) - \beta^3 I_h (1 + \cos(\beta l) \cosh(\beta l))) \quad (24)$$

The individual numerator functions are given below.

$$N_{\theta, V_a}(s) = -\beta C_a \rho A \times ((\cos(\beta l) + \cosh(\beta l)) (\cos(\beta (l - x_1)) - \cos(\beta (l - x_2)) + \cosh(\beta (l - x_1)) - \cosh(\beta (l - x_2))) + (\sin(\beta l) + \sinh(\beta l)) (\sin(\beta (l - x_1)) - \sin(\beta (l - x_2)) - \sinh(\beta (l - x_1)) + \sinh(\beta (l - x_2)))) \quad (25)$$

$$N_{V_s, T}(s) = -\frac{C_a}{C_s} N_{\theta, V_a}(s) \quad (26)$$

$$N_{Y, V_a}(s) = C_a (2 \rho A ((-\cosh(\beta x_1) + \cosh(\beta x_2)) \sin(\beta l) + (-\cos(\beta x_1) + \cos(\beta x_2)) \sinh(\beta l)) + \beta^3 I_h (\cos(\beta (l - x_1)) - \cos(\beta (l - x_2)) + (\cos(\beta x_1) - \cos(\beta x_2)) \cosh(\beta l) - \cosh(\beta (l - x_1)) + \cosh(\beta (l - x_2)) + \cos(\beta l) (-\cosh(\beta x_1) + \cosh(\beta x_2)) + (\sin(\beta x_1) - \sin(\beta x_2)) \sinh(\beta l) + \sin(\beta l) (\sinh(\beta x_1) - \sinh(\beta x_2)))) \quad (27)$$

$$\begin{aligned}
N_{V_s, V_s} = & \frac{1}{2} \beta C_a C_s \times \\
& (\rho A((-4 \cos(\beta l) - \cos(\beta(l-2x_1)) - \cos(\beta(l-2x_2)) - 2(\cos(\beta(l-x_1-x_2)) + \\
& \cos(\beta(l+x_1-x_2)))) \cosh(\beta l) + 4(-\cos(\beta x_1) + \cos(\beta x_2)) \cosh(\beta x_1) + \\
& \cos(\beta l)(-\cosh(\beta(l-2x_1)) - \cosh(\beta(l-2x_2)) + 2(\cosh(\beta(l-x_1-x_2)) + \\
& \cosh(\beta(l+x_1-x_2)))) + 4(\cos(\beta x_1) - \cos(\beta x_2)) \cosh(\beta x_2) + (-\sin(\beta(l-2x_1)) - \\
& \sin(\beta(l-2x_2)) + 2(\sin(\beta(l-x_1-x_2)) + \sin(\beta(l+x_1-x_2)))) \sinh(\beta l) + \\
& \sin(\beta l)(\sinh(\beta(l-2x_1)) + \sinh(\beta(l-2x_2)) + 2(-\sinh(\beta(l-x_1-x_2)) - \\
& \sinh(\beta(l+x_1-x_2)))) + \\
& \beta^3 I_h ((\cosh(\beta(l-2x_1)) + \cosh(\beta(l-2x_2)) - 2 \cosh(\beta(l-x_1-x_2))) \sin(\beta l) - \\
& 2 \cosh(\beta x_2) \sin(\beta x_1) + 2 \cosh(\beta l)(\sin(\beta l) - \sin(\beta(l+x_1-x_2))) + \\
& 2 \cosh(\beta x_1)(\sin(\beta x_1) - \sin(\beta x_2)) + 2 \cosh(\beta x_2) \sin(\beta x_2) + (\cos(\beta(l-2x_1)) + \\
& \cos(\beta(l-2x_2)) - 2 \cos(\beta(l-x_1-x_2))) \sinh(\beta l) - 2 \cos(\beta x_2) \sinh(\beta x_1) + \\
& 2(-(\cosh(\beta(l-x_1)) \sin(\beta(l-x_1))) + \cosh(\beta(l-x_2)) \sin(\beta(l-x_1)) + \\
& \cosh(\beta(l-x_1)) \sin(\beta(l-x_2)) - \cosh(\beta(l-x_2)) \sin(\beta(l-x_2)) - \sin(\beta(x_1-x_2)) - \\
& \cos(\beta(l-x_1)) \sinh(\beta(l-x_1)) + \cos(\beta(l-x_2)) \sinh(\beta(l-x_1)) + \\
& \cos(\beta(l-x_1)) \sinh(\beta(l-x_2)) - \cos(\beta(l-x_2)) \sinh(\beta(l-x_2)) - \sinh(\beta(x_1-x_2))) \\
& + 2 \cos(\beta l)(\sinh(\beta l) - \sinh(\beta(l+x_1-x_2))) + 2 \cos(\beta x_1)(\sinh(\beta x_1) - \\
& \sinh(\beta x_2)) + 2 \cos(\beta x_2) \sinh(\beta x_2)))
\end{aligned} \tag{28}$$

The duality of $N_{\theta, V_s}(s)$ and $N_{V_s, T}(s)$ which is expressed in (26) has interesting physical implications. Note that the procedure used here also yields the previously known numerator functions $N_{\theta, T}(s) = 2\beta\rho A(1 + \cos(\beta l) \cosh(\beta l))$ and $N_{V, T}(s) = l\rho A[\sin(\beta l) + \sinh(\beta l)]$ in agreement with [8].

4 Rational Laplace Domain Transfer Functions

The transfer functions presented above are exact relationships (for the model used) in terms of transcendental functions of β . The desired form of the transfer functions (23) is in terms of ratios of polynomials in the Laplace operator s . For this purpose, the numerator and denominator transcendental functions can be rationalized as shown by Schmitz [8] using the Maclaurin series expansion,

$$f(\beta) = \frac{\beta^p}{p!} \frac{d^p f(0)}{d\beta^p} (1 + k_1\beta^4 + k_2\beta^8 + \dots) \tag{29}$$

or in equivalent product form

$$f(\beta) = \frac{\beta^p}{p!} \frac{d^p f(0)}{d\beta^p} \prod_{i=1}^N \left[1 - \left(\frac{\beta}{\beta_i} \right)^4 \right] \tag{30}$$

The derivative term in the leading coefficient $\frac{d^p f(0)}{d\beta^p}$ is obtained by differentiating the function $f(\beta)$ until a non-zero constant is obtained. The p is the index of that first non-zero derivative and N corresponds to the number of modes to be retained in the rationalized representation. This procedure leads to the following rational expressions for the elements of (23) in terms of s .

$$G_{V_s, V_s}(s) = \frac{C_a C_s (x_1 - x_2)}{EI} \prod_{i=1}^N \frac{((s/\gamma_{vi})^2 + 1)}{((s/\Omega_i)^2 + 1)} \tag{31}$$

$$G_{V_s, T}(s) = -\frac{C_s \rho A}{24 I_T E I} (4l ((l - z_1)^3 - (l - z_2)^3) - ((l - z_1)^4 - (l - z_2)^4)) \prod_{i=1}^N \frac{((s/\gamma_{T_i})^2 + 1)}{((s/\Omega_i)^2 + 1)} \quad (32)$$

$$G_{\theta, V_s}(s) = -\frac{C_s}{C_s} G_{V_s, T}(s) \quad (33)$$

$$G_{y, T}(s) = \frac{l}{s^2 I_T} \prod_{i=1}^N \frac{((s/\lambda_{T_i})^2 - 1)}{((s/\Omega_i)^2 + 1)} \quad (34)$$

$$G_{\theta, T}(s) = \frac{1}{s^2 I_T} \prod_{i=1}^N \frac{((s/\omega_{T_i})^2 + 1)}{((s/\Omega_i)^2 + 1)} \quad (35)$$

$$G_{y, V_s}(s) = \frac{C_s}{24 I_T E I} (12 I_h (-z_1 + z_2)(x_1 + x_2 - 2l) + l \rho A (2l^2(z_1^2 - z_2^2) - z_1^4 + z_2^4)) \prod_{i=1}^N \frac{((s/\lambda_{V_i})^2 - 1)}{((s/\Omega_i)^2 + 1)} \quad (36)$$

where the total rotary inertia I_T is defined $I_T \triangleq I_h + \frac{l^3 \rho A}{3}$.

The transcendental equations (24)–(29) which form the various transfer functions in (23) may seem too complex to be of practical utility, but it is the finite truncation (31)–(36) of these transcendental functions that is of ultimate interest for control purposes. To get an N mode finite truncation first N roots of these transcendental equations must be obtained numerically, and once numerical methods are resorted to the complexity of the transcendental equations is really not an issue as long as it is practical to obtain the roots of these equations.

5 The Experimental Beam

The transfer functions were verified using a beam experiment at the US Air Force Academy, Frank J. Seiler Research Laboratory. The experiment consists of a 2 m long uniform aluminum beam of rectangular cross section (76.2 mm $\times$ 6.35 mm), hanging vertically from a very low friction hinge. The hinge is in a knife edge arrangement as illustrated in Figure 3, and is situated in such a way that the root of the flexible beam is coincident with the hinge axis. This also allows PZT actuator/sensor pairs to be located with one edge coincident with the hinge axis. The hinge arrangement has effectively zero inertia, however, a hub inertia has been added in order to increase pole-zero separation. Although the system has no torque motor, for purposes of modal testing, torque inputs are achieved by applying impulsive force to a rigid moment arm extending from the hinge axis. A PCB Piezotronics force hammer is used to provide the force. Because the angular displacements of the hinged joint are very small during modal testing, it was possible to get an adequate measurement of hub angle by measuring translational displacement of a point on the arm extending from the hinge. A Kaman KD-2300-10CU non-contact displacement measuring system was used for this measurement. This sensor has a rated 3dB frequency response range of static to 50 kHz., a measurement range of 25.4 mm, and 0.254 mm midrange resolution. The piezoelectric ceramic elements are Vernitron PZT5A patches mounted in pairs on opposite faces of the beam. A 63.5 mm long by 38.1 mm wide piezoelectric material

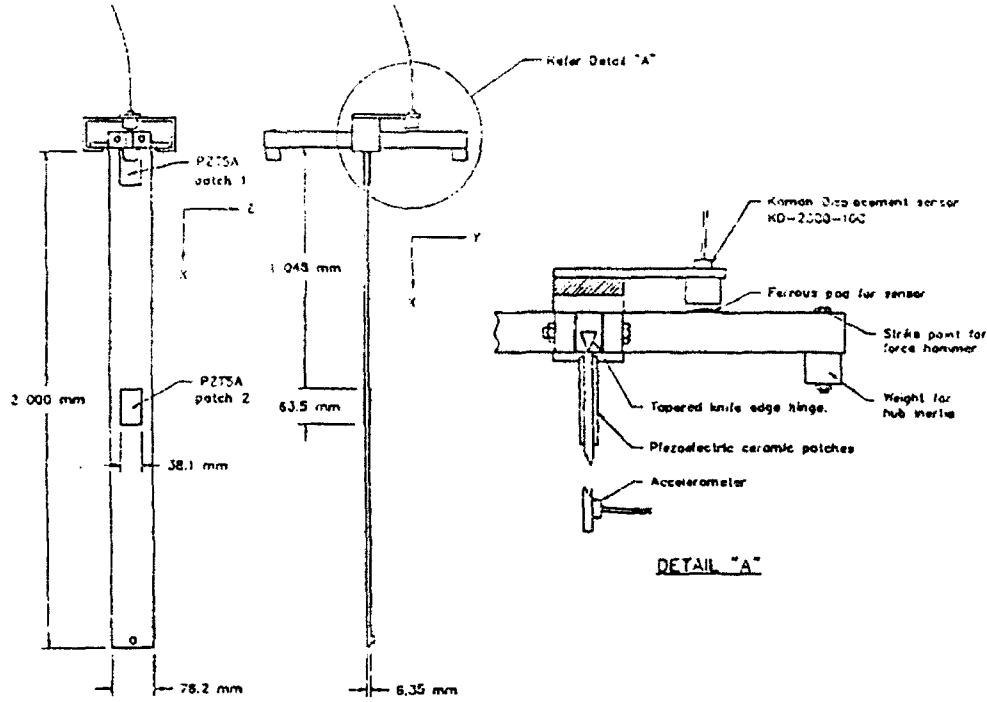


Figure 3: The Experimental Beam

patch is attached to each side of the root of the beam so $x_1 = 0$, $x_2 = 63.5$ mm, and $w = 38.1$ mm. A second pair is located roughly midway down the beam at $x_1 = 1.048$ m, $x_2 = 1.111$ m. The PZT patches used for both actuation and sensing are identical, and hence $t_a = t_s$.

Unlike the idealized sensing relationship given in (6), a more precise model of the PZT sensing behavior includes an RC high-pass filter characteristic which can be expressed (in the Laplace domain) as follows:

$$V_s(s) = C_s \left(\frac{s R_m C_t}{s R_m C_t + 1} \right) \frac{\partial y(x, s)}{\partial x} \Big|_{x_1}^{x_2} \quad (37)$$

where R_m is the resistance of the measurement process (data acquisition system or oscilloscope) and C_t is the effective total capacitance of the PZT patch (as opposed to C in (6), which is capacitance per unit area). In order to force the sensor response to behave as (6) in the frequency range of interest (i.e., flat frequency response) it is desirable to introduce a large measurement impedance R_m . For this purpose, the output of the PZT sensors is processed through a high impedance ($10^{12} \Omega$) unity gain buffer amplifier. This moves the high-pass break frequency $\frac{1}{R_m C_t}$ to well below 1 Hz. Beam tip motion is measured using a PCB Piezotronics Model 336A02 accelerometer. The rated $\pm 5\%$ amplitude sensitivity deviation range for the accelerometer is 1-2000 Hz. A Tektronix model 2630 Fourier Analyzer is used for modal testing. The parameters for experimental system are tabulated in Table 1. Table 2 contains the parameters of the piezoelectric material.

Hub inertia, I_h	0.0348 N-m-s ²	Beam length, l	2.0 m
Volumetric mass density, ρ	2712.6 Kg/m ³	Cross sectional area, A	483.87×10^{-6} m ²
Young's Modulus, E	69.0×10^9 N/m ²	Area moment of inertia, I	1.63×10^{-9} m ⁴

Table 1: Experimental Beam Parameters

Charge Constant, d_{31}	-171×10^{-12} m/v	Voltage Constant, g_{31}	-11.4×10^{-3} Vm/N
Coupling Coefficient, k_{31}	-0.340	Capacitance, C	68.35 μ F/m ²
Surface area, $w \times (x_2 - x_1)$	0.0024 m ²	Thickness t_a	3.05×10^{-4} m
Young's Modulus, E_a	69.0×10^9 N/m ²		

Table 2: PZT5A Parameters

Roots for the various transcendental equations (24)–(29), needed to model the system, are given in Table 3. These roots, which give a theoretical estimate of the system poles and zeros, are obtained for the experimental beam for which the parameters have been listed in Table 2. Note that the system poles and zeros do not depend on the particular piezoelectric element chosen—they depend only on the location and size of the element (in the case of the zeros) and the beam parameters. The transfer function gains are dependent on the particular piezoelectric material as can be seen from the relations (31)–(36). Upon consideration of the various poles and zeros in Table 3 some interesting observations can be made regarding the system from a controls point of view. The transfer functions:

- $G_{\theta,T}(s)$, $G_{V_{a1},V_{a1}}(s)$, and $G_{V_{a2},V_{a2}}(s)$ have colocated and compatible actuators and sensors, and accordingly have only minimum phase zeros, [3], [8], and exhibit the well known pole/zero interlacing property [9].
- $G_{V_{a2},V_{a2}}(s)$ has an exact pole/zero cancellation at 27.55 Hz (the third mode), due to actuator/sensor location. Thus, that mode is uncontrollable using this colocated piezoelectric actuator/sensor pair. The first mode of $G_{V_{a1},V_{a1}}(s)$ also has a near pole/zero cancellation.
- $G_{\theta,V_{a1}}(s)$ and $G_{V_{a1},T}(s)$ exhibit the well-known pole-zero interlacing property [9] characteristic to

mode	$\Delta(s)$ (Ω_i)	$N_{\theta,T}(s)$ (ω_{T_i})	$N_{\theta,V_a}(s)$ (ω_{v_i})	$N_{y,T}(s)$ (λ_{T_i})	$N_{y,V_a}(s)$ (λ_{v_i})	$N_{V_{a1},V_{a1}}(s)$ (γ_{v1_i})	$N_{V_{a2},V_{a2}}(s)$ (γ_{v2_i})	$N_{V_{a1},T}(s)$ (γ_{T_i})
1	$\pm j5.477$	$\pm j1.313$	$\pm j5.851$	± 4.176	± 5.544	$\pm j5.487$	$\pm j5.624$	$\pm j5.851$
2	$\pm j15.33$	$\pm j8.226$	$\pm j18.97$	± 22.57	± 24.06	$\pm j15.84$	$\pm j15.595$	$\pm j18.97$
3	$\pm j27.55$	$\pm j23.03$	$\pm j39.59$	± 55.73	± 58.29	$\pm j30.08$	$\pm j27.55$	$\pm j39.59$
4	$\pm j47.31$	$\pm j45.14$	$\pm j67.73$	± 103.6	± 108.1	$\pm j51.622$	$\pm j48.68$	$\pm j67.73$
5	$\pm j75.84$	$\pm j74.61$	$\pm j103.4$	± 166.3	± 173.5	$\pm j81.91$	$\pm j76.46$	$\pm j103.4$

Table 3: Roots of Various Transcendental Equations in Hz

transfer functions for systems with colocated actuator/sensor pairs. This is somewhat unexpected, as one might anticipate some non-minimum phase zeros due to the fact that x_2 is not coincident with the hinge point. Plotting $N_{\theta,V_{a1}}(s)$ and $N_{V_{a1},T}(s)$ for real values of s in the interval $\{0, 35000\text{Hz}\}$ reveals no zero crossings in that interval, which supports the conclusion of no non-minimum phase zeros, however this conclusion would be difficult to verify experimentally.

- $G_{y,T}(s)$, having noncolocated actuators and sensors at the extreme ends of the flexible beam, has only non-minimum phase zeros, as is well-known [8].
- $G_{y,V_{a1}}(s)$ has only non-minimum phase zeros in the interval of primary interest $\{0, 100\text{Hz}\}$. However, analytically, minimum phase zeros are predicted at 1,935 Hz, 11,205 Hz, and 27,641 Hz, for s in the real interval $\{0, 35000\text{Hz}\}$. Experimentally, zeros can be detected at about 1500 Hz, and 8,150 Hz, however accurate experimental data was difficult to obtain at these frequencies. These zeros are well beyond the frequency range for which the Bernoulli-Euler assumptions² are considered accurate. Note that 1,935 Hz is beyond the 20th mode. The existence of these zeros, while quite interesting, is not entirely unexpected, since x_2 is not colocated with the hinge point.
- $G_{V_{a1},V_{a1}}(s)$ and $G_{V_{a2},V_{a2}}(s)$ have closely spaced poles and zeros, which suggests that these colocated transfer functions will allow very little control authority over the oscillation modes. When I_h is increased, the separation increases, particularly at the higher frequencies, thus higher modes become more controllable as I_h is increased.

6 Experimental Results

The results of the experiments conducted to verify the various transfer functions are presented as Bode magnitude plots. The measured values and the theoretical values, for a particular transfer function, are both plotted on the same graph. The results for the tip position transfer functions $G_{y,T}(s)$ and $G_{y,V_a}(s)$ (Figure 6) are in fact presented in terms of the measured output variable, tip acceleration. From the plots, the accuracy of the theoretical predictions can be easily verified. A damping ratio of 0.2% is added to each complex conjugate pair of poles and zeros in the theoretical plots to approximate material damping. For this three input (T , V_{a1} , and V_{a2}) and four output (θ , V_{s1} , V_{s2} , and Y_{tip}) system there are actually twelve transfer functions of interest. Here we present plots for only the seven transfer functions which have either all minimum phase zeros, or all non-minimum phase zeros. Noncolocated actuator/sensor pairs, with at least one actuator or sensor located away from the beam's endpoints, can yield transfer functions which exhibit both types of zeros. Preliminary results for these special cases are presented in the subsection which

²Generally considered applicable for the first 10 modes of long slender beams.

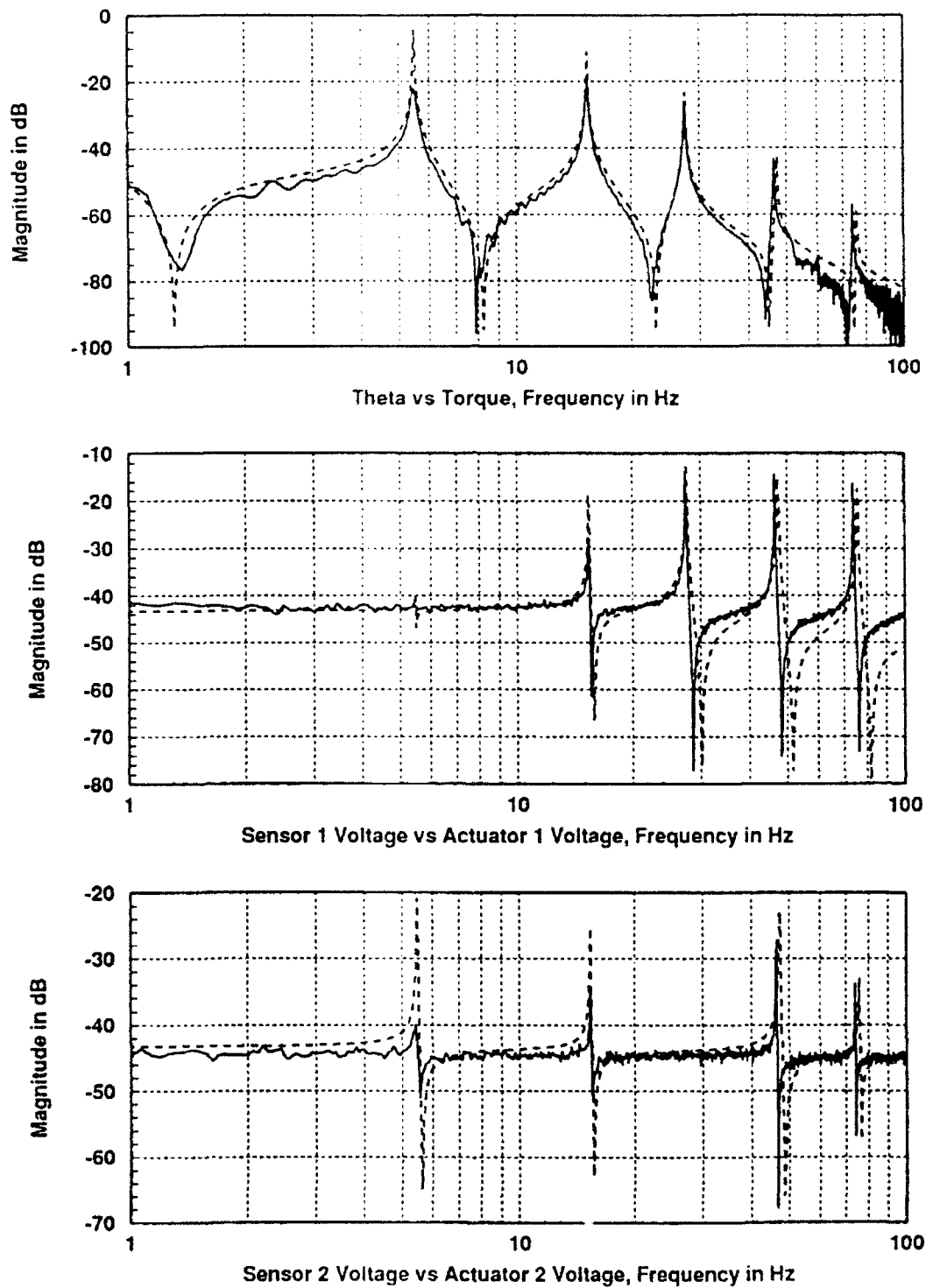


Figure 4: Colocated Transfer Functions $G_{\theta,T}(s)$, $G_{V_{s1},V_{a1}}(s)$, and $G_{V_{s2},V_{a2}}(s)$. Solid—Experiment, Dashed—Theory

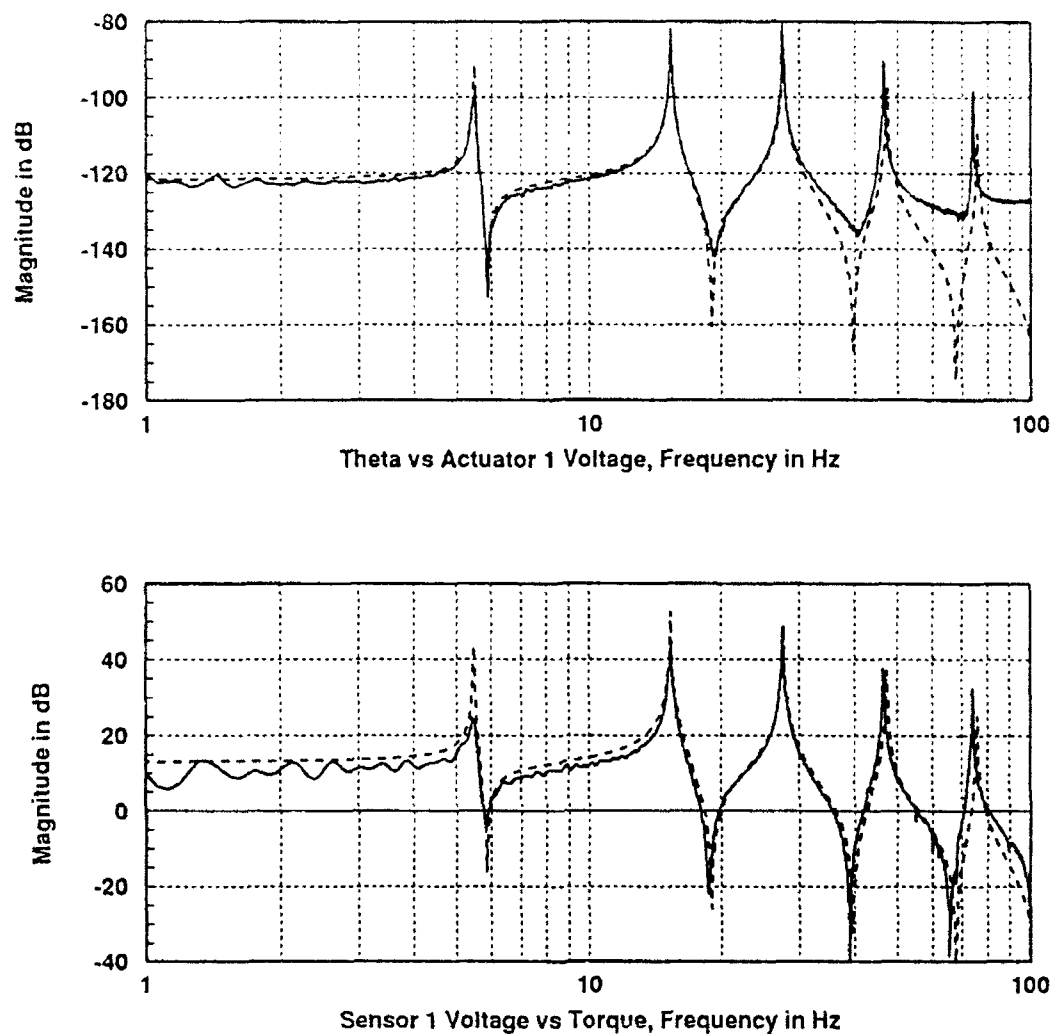


Figure 5: Minimum Phase Noncolocated Transfer Functions $G_{\theta, V_{a1}}(s)$ and $G_{V_{s1}, T}(s)$. Solid—Experiment, Dashed —Theory

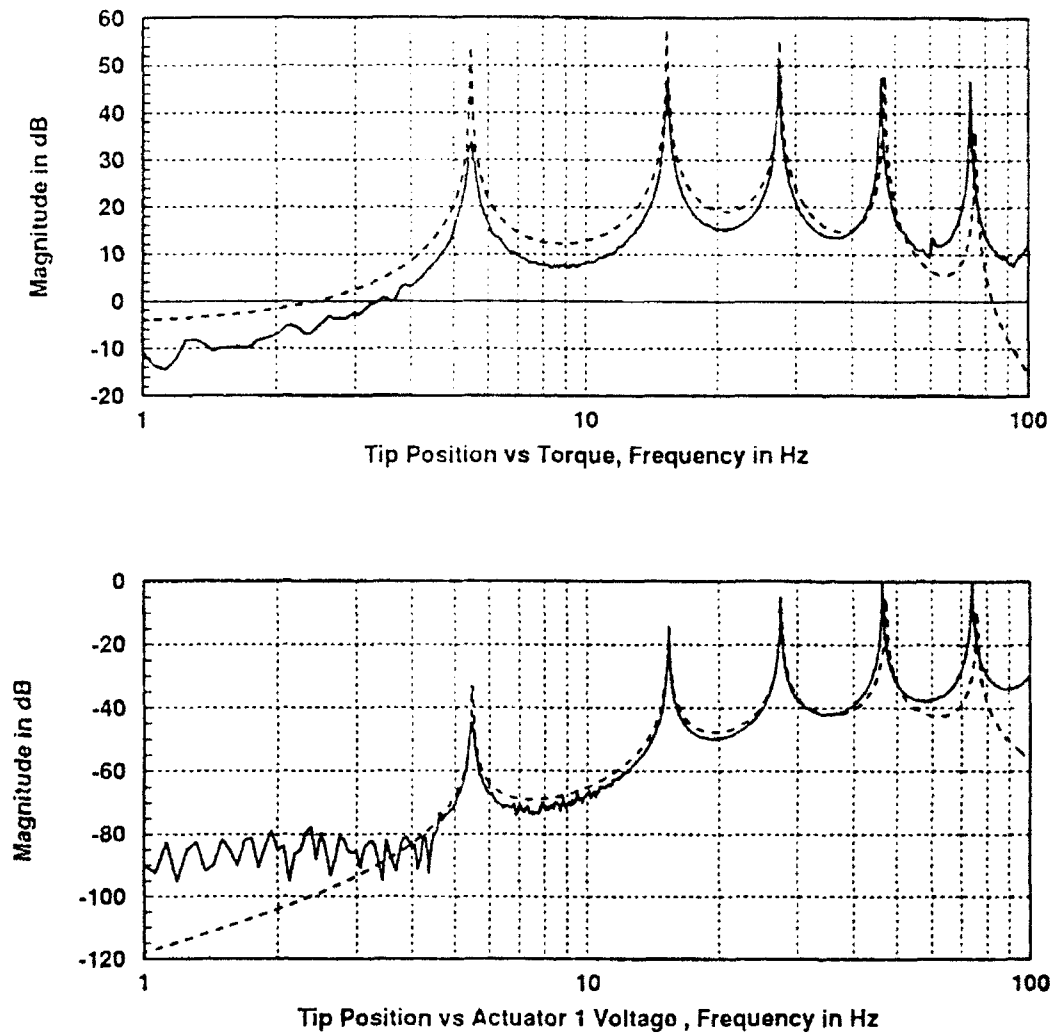


Figure 6: Non-minimum Phase Noncolocated Transfer Functions $G_{y,T}(s)$ and $G_{y,v_{a1}}(s)$. Solid—Experiment, Dashed —Theory

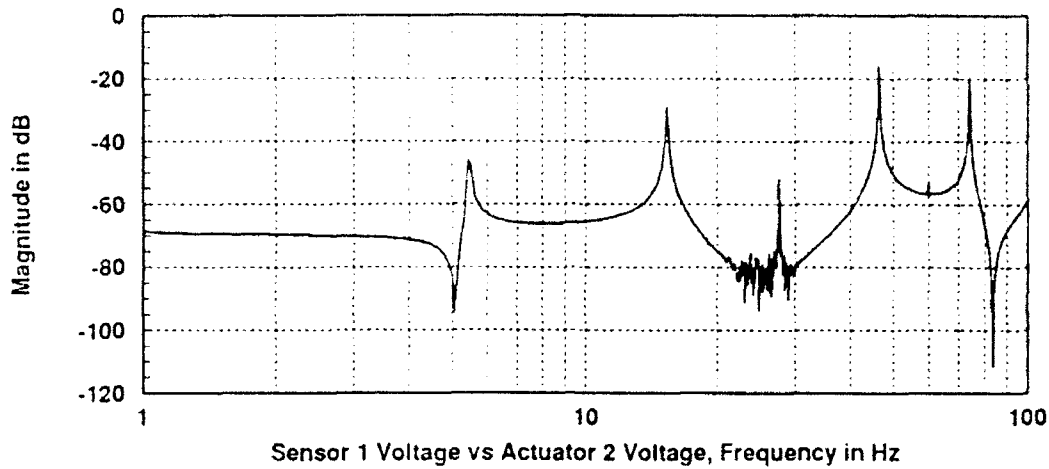


Figure 7: Noncolocated Transfer Function with Minimum Phase and Non-minimum Phase zeros, $G_{V_{s1}, V_{a2}}(s)$.

follows. For all of the plots, the predicted pole and zero frequencies are seen to agree well with experiment. The third mode pole/zero cancellation in $G_{V_{s2}, V_{a2}}(s)$ and the first mode near cancellation in $G_{V_{s1}, V_{a1}}(s)$ are also verified by the experimental results. Non-minimum phase zeros cannot be observed directly as peaks or valleys in frequency response plots, however they do have a substantial influence on the curve shapes. The general agreement between the shapes of the corresponding analytical and experimental curves provides evidence of the accuracy of the predicted non-minimum phase zeros. The experimental data (Figures 4a and 4b) for the transfer functions $G_{\theta, T}(s)$ and $G_{\theta, V_a}(s)$ becomes noisy at higher frequencies so that the zeros appear truncated. This is due to the resolution limitations of the Kaman sensor employed. Similarly, for the transfer functions $G_{y, T}(s)$ and $G_{y, V_a}(s)$ (Figure 6) the data quality is reduced somewhat at the lower end of the frequency range, due to weak response of the accelerometer at low frequencies. Some difficulty was encountered in obtaining data for which the predicted and measured magnitudes matched well for the colocated voltage to voltage transfer functions (Figures 4b and 4c). The data presented represents the best match in magnitude achieved, however it should be noted that varying the input voltage amplitude by a factor of 100 can cause frequency response plot magnitude variations of as much as 7 dB. The explanation for this uncertainty is unknown to the authors. The data presented uses a peak input amplitude of 60 volts.

6.1 Transfer Functions with Minimum Phase and Non-minimum Phase Zeros

Colocated and compatible actuator/sensor pairs, always yield minimum phase transfer functions with pole/zero alternation. For the system we have considered, when noncolocated actuator sensor pairs are located at the beam's extreme endpoints, the transfer functions have only non-minimum phase zeros. Noncolocated actuator/sensor pairs, with at least one actuator or sensor located away from the beam's endpoints, can yield transfer functions which exhibit both types of zeros. Experimental validation of this effect has been ob-

tained. Here we present experimental results only the $G_{V_{11},V_{22}}(s)$ due to space limitations. As of the present time, the analytical results have not been prepared for comparison, however, the poles and zeros have been computed and are in excellent agreement with the experimental results obtained.

7 Optimal Passive Damping Design

A long range goal of this work is to ultimately develop a simultaneously optimal active and passive damping design for the experimental beam system. The active damping will incorporate closed-loop feedback control of selected actuator/sensor pairs. The passive damping measure will be a constrained viscoelastic layer surface treatment. A procedure has been developed [11] whereby the active and passive damping designs can be simultaneously optimized. Based upon the models developed in this report it will be possible to carry out this procedure. As a first step toward this goal, a damping treatment was designed which is optimized to damp the 3rd vibration mode of the experimental beam. The passive damping optimization procedure involves selecting a constraining layer section length which optimizes the damping at a designer prescribed frequency. A method developed by Plukett and Lee [12] has been employed for the design and analysis. This beam will be used as a suboptimal design case for comparison to the simultaneously optimal design. The beam also serves to provide data needed to evaluate our ability to accurately predict the damping achieved by the treatment. This predictive ability will be crucial for the simultaneously optimal design.

The materials selected for the prototype damping treatment include, mild steel shim stock (0.012 in thick) for the constraining layer, and 10 mil Scotchdamp³ SJ-2015X Type 1010 Viscoelastic Polymer. Steel was chosen for the constraining layer, due to its relatively high stiffness, low cost, well known mechanical properties, and ease of workability. For space applications, where weight is of concern, graphite fiber layers may be used very effectively. The 10 mil thickness Scotchdamp rendered optimal constraining layer sections of practical length for easy construction. An optimal section length of 3.65 inches for the constraining layer was obtained for the targeted third mode. Mechanical data for the viscoelastic polymer was extrapolated from a nomogram provided by 3M, for a temperature of 18°C⁴. Experimental damping ratios were evaluated from the $G_{y,T}(s)$ frequency response data (Figure 8) for the damped system. The half power method was used to compute the damping ratios. The theoretical data resulted in damping coefficients on the average of 40% lower than experimental values. Computational errors on the order of as much as 100% are not unusual for passive damping analysis. The authors have no explanation for the higher actual damping coefficient, but it is certainly better than getting less than the predicted value of damping. The treatment just worked too well. Usually, accurate predictions are primarily a matter of having the proper viscoelastic material property data. It is our intention to seek updated Scotchdamp property data to permit improved predictive

³The viscoelastic polymer materials used in these experiments were contributed by 3M Corporation.

⁴Unfortunately, this is about as warm as the lab ever gets.

Mode Index	1	2	3	4	5
Mode Frequency (Undamped) Hz	5.188	15.31	28.25	47.25	75.75
Experimental Damping Ratio	0.071	0.070	0.062	0.047	0.056
Predicted Damping Ratio	0.050	0.037	0.032	0.027	0.023

Table 4: Experimental and Predicted Damping Results

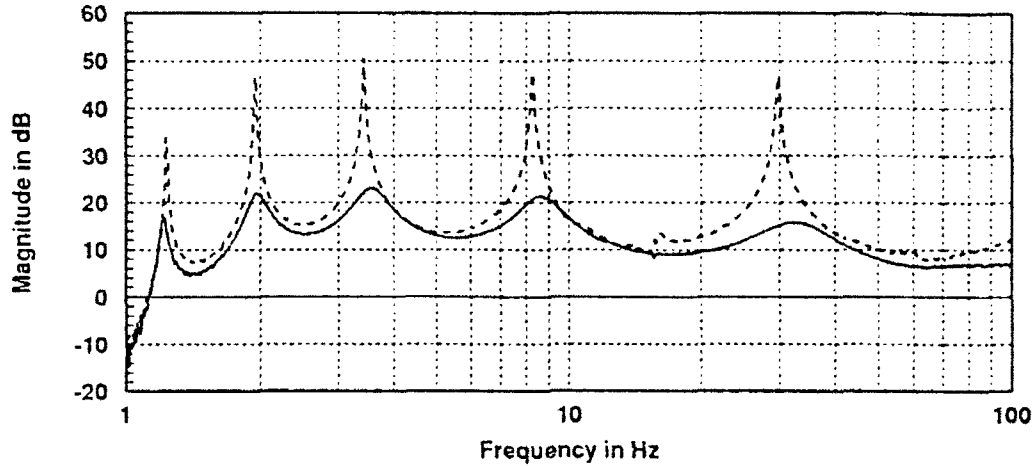


Figure 8: Comparison of Damped and Undamped $G_{y,T}(s)$. Solid—Undamped, Dashed—Damped

abilities. It is interesting to note that the natural frequencies of the beam were changed very little by the addition of the damping treatment. One might expect the natural frequencies to increase due to the added stiffness, but the added inertia counteracts this effect.

8 Conclusion

A method to obtain exact transfer functions for a flexible slewing piezoelectric laminate beam has been presented. The resulting transcendental forms of transfer functions represent multi-input, multi-output pair combinations for the slewing flexible beam. The transfer functions have been reduced to rational form using Maclaurin series expansions as in [3]. These transfer functions have been experimentally verified using the beam experiment at the US Air Force Frank J. Seiler Research Laboratory. The exact transfer functions, derived based on certain simplifying assumptions, can be obtained with relative ease, and the experimental verification of the derived transfer functions indicates that the presented modelling method along with the resulting transfer functions can be used to design controllers for active and passive vibration damping. Preliminary passive damping results for the experimental system show the effectiveness of the proposed passive damping treatment, however it will be necessary to improve our analytical damping prediction capabilities before the simultaneously optimal active/passive design can be applied effectively.

References

- [1] H. R. Pota and T. E. Alberts, "Multivariable Transfer Functions for a Slewing Piezoelectric Laminate Beam," Proc. IEEE Int. Conf. on Systems Engineering, Sept 17-19, Kobe, Japan, 1992.
- [2] C.-K. Lee, "Piezoelectric Laminates for Torsional and Bending Modal Control: Theory and Experiment," Ph.D. dissertation, Cornell University, 1987.
- [3] T. E. Alberts and J. A. Colvin, "Observations on the Nature of Transfer Functions for Control of Piezoelectric Laminates," J. of Intelligent Material Systems & Structures, Vol 2, No 4, pp 528-541, Oct 1991.
- [4] T. Bailey and J. E. Hubbard, "Distributed Piezoelectric-Polymer Active Vibration Control of a Cantilever Beam," J. Guidance, Control and Dynamics, Vol 8, No 5, pp 605-611, 1985.
- [5] A. Baz and S. Poh, "Performance of an Active Control System with Piezoelectric Actuators," J of Sound and Vibrations, Vol 126, No 2, PP 327-343, 1988.
- [6] J. Pan, C. H. Hansen, and S. D. Snyder, "A Study of the Response of a Simply Supported Beam to Excitation by a Piezoelectric Actuator," Proceedings of the Conference on Recent Advances in Active Control of Sound and Vibration, VPI & SU, Virginia, pp 39-49, April 1991.
- [7] E. Garcia, D. Inman, and J. Dosch, "Vibration Suppression using Smart Structures," ASME 1991 Winter Meeting, AD-Vol 24, Smart Structures and Materials, pp 167-172, 1991.
- [8] E. Schmitz, "Experiments on the End-Point Position Control of A Very Flexible One-Link Manipulator," Ph.D. Dissertation, Stanford University, Stanford CA., 1985.
- [9] W. Gevarter, "Basic Relations for Control of Flexible Vehicles," AIAA J., Vol. 8, No. 4, pp 666-672, April 1970.
- [10] A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill, 1962.
- [11] T.E. Alberts and Joseph V. Harrell, "Optimal Passive/Active Control Design Synergism for Flexible Structures," Submitted to the AIAA Journal of Guidance, Control and Dynamics, July 12, 1991.
- [12] R. Plunkett and C.T. Lee, "Length Optimization for Constrained Viscoelastic Layer Damping," J. Accoust. Soc. Amer., Vol. 48, No. 1, 1970, pp. 150-161.

**VELOCITY AND VORTICITY MEASUREMENTS
IN
TRANSIENT OSCILLATORY SEPARATING BOUNDARY LAYER FLOWS**

**B. Terry Beck
Associate Professor
Department of Mechanical Engineering**

**Paul K. Berg
Graduate Student
Department of Mechanical Engineering**

**Kansas State University
Durland Hall
Manhattan, KS 66506**

**Final Report for:
Summer Faculty Research Program/
Graduate Student Research Program
Frank J. Seiler Research Laboratory
US Air Force Academy**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

September 1992

**VORTICITY AND VELOCITY MEASUREMENTS
IN
TRANSIENT OSCILLATORY SEPARATING BOUNDARY LAYER FLOWS**

B. Terry Beck
Associate Professor

Paul K. Berg
Graduate Student

Department of Mechanical Engineering
Kansas State University

Abstract

The velocity and vorticity distribution within a transient oscillatory separating boundary layer was investigated using a single-component Laser Doppler Velocimeter System. The flow was initiated above a flat plate test model by means of a computer-controlled rotating spoiler (flap), mounted above the model surface. The tests were conducted in a water tunnel test facility, and dye injection was also utilized for visualization of the flow separation phenomena. The rotating spoiler subjected the plate below to a time-dependent spatial pressure gradient, inducing periodic flow separation and vortex shedding from the region near the plate and downstream of the spoiler. Measurements of both horizontal and vertical velocity components were made by rotating the optics of the LDV system. These profile measurements were obtained for discrete angular flap positions, thus mapping out the spatial and time-dependent flow field downstream of the flap. From the separate velocity component profiles, a computerized scanning algorithm was implemented to obtain both scan-averaged velocity and velocity gradient fields. Using this technique resulted in remarkably smooth results, in spite of the limited spatial resolution of the transient measurements. Clear evidence of reverse flows and flow bifurcation is indicated from the measurements near the region of boundary layer separation. The effect of flap frequency on the separation phenomena was also investigated.

VORTICITY AND VELOCITY MEASUREMENTS IN TRANSIENT OSCILLATORY SEPARATING BOUNDARY LAYER FLOWS

**B. Terry Beck
Paul K. Berg**

I. INTRODUCTION

Areas of active research in fluid mechanics presently include unsteady flows, separating flows, and dynamic stall. In each of these flows, the vorticity is a very useful parameter. In dynamic stall, the central focus of research has been the characterization of the movement and shedding of the bound vortices.

In the study of unsteady separating flows, Reynolds and Carr (1985) have suggested that understanding the transport of vorticity is the key to understanding the mechanisms that control these flows. In an adverse pressure gradient, the wall becomes a source for vorticity to diffuse into the boundary layer. The roll-up and shedding of these vortices has been the subject of much research. In general, however, these studies have been largely qualitative in nature due to the difficulty of measuring vorticity.

The motivation for the current research effort relates to the unsteady boundary layer separation on an airfoil undergoing rapid changes in angle of attack. It has been shown that under such circumstances the angle of attack associated with a rapidly pitched airfoil is extended well beyond the stall condition for a static airfoil. After separation occurs, the separation point moves upstream toward the leading edge, resulting in the formation of a leading edge vortex. This dynamic vortex is subsequently shed from the airfoil carried away downstream. Initiation of boundary layer separation in the presence of an adverse pressure gradient is of considerable importance to the dynamic stall phenomena, and to boundary layer control (Lovato, 1992).

Recent research efforts have focussed specific attention on the formation (on-set) of flow separation in transient boundary layer flows, in an effort to enhance understanding of the physical mechanisms associated with this phenomena, and to identify the governing parameters and the appropriate separation criteria. Detailed local measurements of the vorticity field may also be key to further developments in this area; in fact, the Scanning LDV Vorticity Measurement System of Beck and Humphreys (1991) was developed specifically for the purpose of providing direct and accurate vorticity field information.

The vanishing of wall shear stress, while adequate to characterize steady-state separation, has been shown to be insufficient for separation in unsteady flows. It has been postulated (Sears, 1956) that unsteady separation should occur where both shear stress and velocity vanish in a frame of reference which moves with the separation point. This is known as

the MRS condition. Measurements of the unsteady separation associated with an impinging jet by Didden and Ho (1985) appear to support this criteria.

II.OBJECTIVES

The primary objective of this research was the investigation of the velocity and vorticity field associated with a transient oscillatory separating boundary layer flow. Flow separation was induced by means of a rotating spoiler (flap) attached to a secondary plate above the flat plate test model. Rotation of the flap produced a time-dependent adverse pressure gradient in the flow field downstream of the spoiler, resulting in periodic boundary layer separation on the surface of the test plate. The major tasks to be accomplished were as follows:

1. Construct a flat plate test model, and utilize an existing flat plate test model with rotating spoiler, to induce periodic boundary layer separation over the top surface of the model. This involved establishing computer control of the micro-stepper motor driven flap.
2. Make all necessary modifications and adjustments to the existing two-component scanning LDV system to allow simultaneous measurement of both the horizontal and vertical velocity fields, as well as the spatial distribution of vorticity.
3. Run flow visualization tests with dye injection in the water tunnel test facility. Investigate the effect of free stream velocity above the test model, spoiler flap rotation frequency, and spacing between the plates on the unsteady boundary layer separation.
4. Use the scanning LDV system to map out the spatial and time dependent velocity and vorticity fields, during unsteady boundary layer separation, for the range of test parameters identified during the flow visualization tests.

III. EXPERIMENTAL METHOD

The experimental equipment and instrumentation used in this investigation consisted of a scanning LDV system, a water tunnel test facility, and a modified flat plate test model with rotating spoiler, along with associated electronics and computing hardware for data acquisition and control. These components were modified from existing hardware which had been utilized on previous research involving transient vorticity measurement (Beck and Humphreys, 1991). The modifications were necessary to allow specific investigation of the transient boundary layer separation phenomenon, and to circumvent certain technical problems with the hardware which were identified during the course of this research.

A. LDV System and Water Tunnel Test Facility

A schematic of the scanning LDV system used in this investigation is shown in Figure 1. The system was a single-component (1D), two-beam, Argon Ion laser (5W) based setup, arranged in a back-scatter configuration. It was a rearrangement of the optics associated with a full two-component (2D) system, and incorporated rotatable optics for measurement of both horizontal and vertical velocity components. The aged laser began experiencing multi-modal

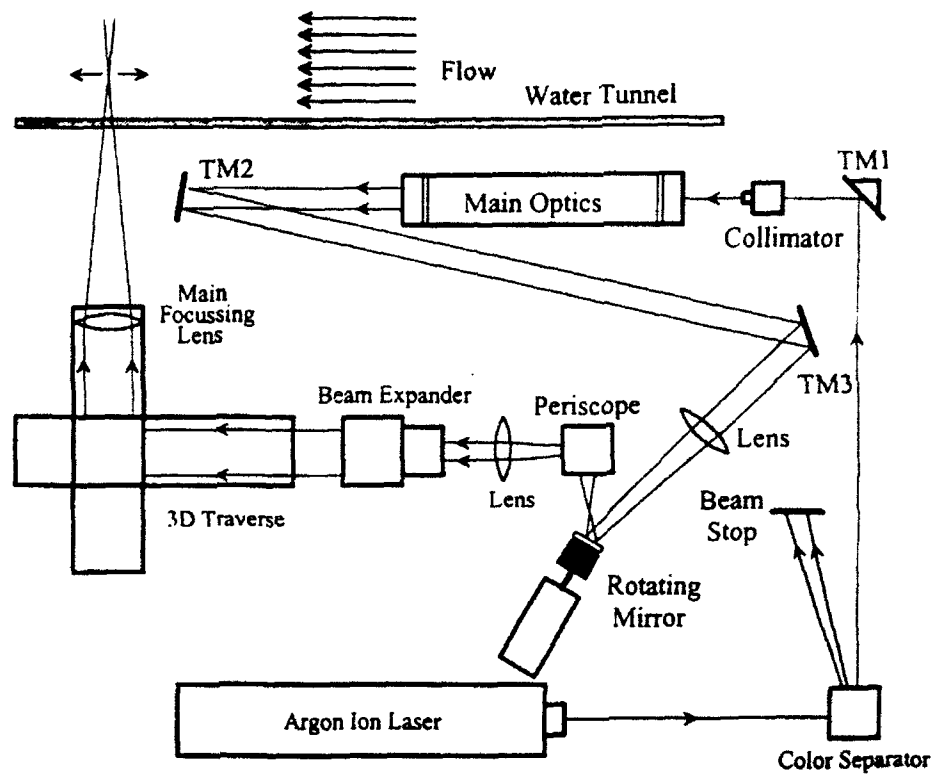


Figure 1: Schematic of Modified LDV System

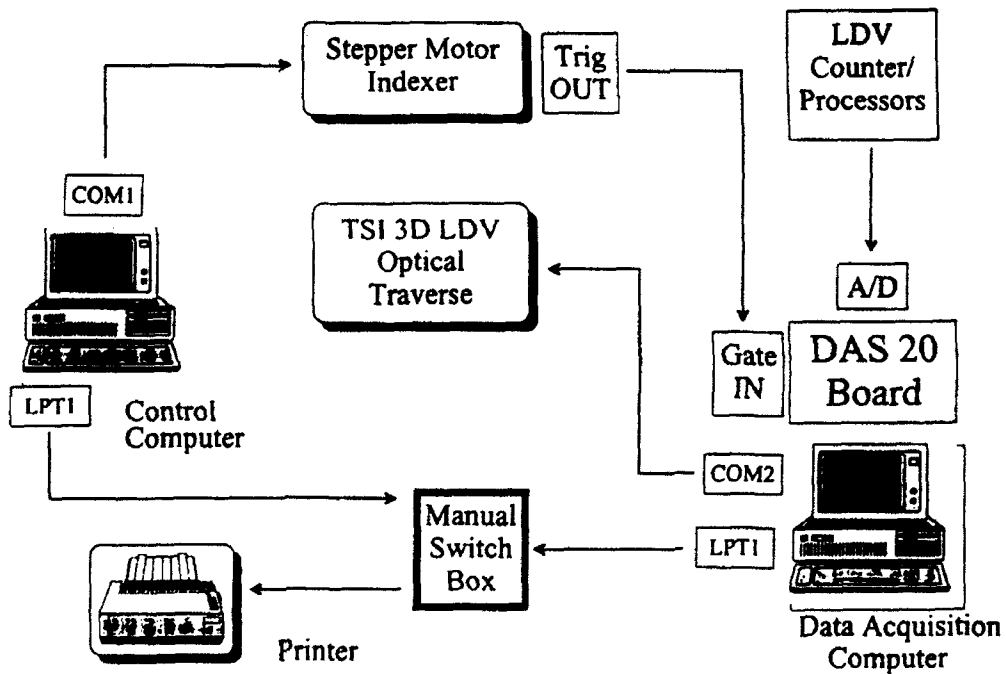


Figure 2: Block Diagram of Data Acquisition and Control

behavior during simultaneous two-color (blue 488.0 nm and green 514.5 nm) operation; however, it was possible to tune the green beam alone to the desired TEM₀₀ mode for single-component velocity measurements. While this required double the number of experiments per test for a full two-component measurement, and manual rotation of the optical axis, it was deemed necessary to avoid the extensive delay associated with repair of the laser.

A TSI Model 9400 3-Axis computer-controlled traverse was used for positioning the measuring volume (beam intersection), providing a range of adjustment of ± 300 mm on all axes and a positioning resolution of ± 0.01 mm. The main focussing lens had a 750 mm focal length, producing a fringe spacing of $4.374 \mu\text{m} \pm$ about 1% for the horizontal velocity component measurements. The system incorporated circular scanning for direct vorticity measurement, as described by Beck and Humphreys (1991); however, due to the time constraints imposed by single-component LDV operation, an alternative post-processing algorithm was developed for determination of vorticity. The focussing lenses, rotating mirror, and periscope, associated with the scanning capability, also provided fine adjustment of optical axis position and beam crossing angle. For the vertical velocity component measurements, these lenses were modified to reduce the beam crossing angle and allow a closer wall approach (without beam obstruction by the test model). The fringe spacing for vertical velocity component measurements was $7.480 \mu\text{m} \pm$ about 1%. Bragg cell frequency shifting was utilized to sense flow direction.

The LDV system was located adjacent to an Eidetics Model 1520 Water Tunnel Test Facility, within the Aeronautics Laboratory at the USAF Academy. The tunnel had a test section opening about 160 cm long, 38 cm wide, and 50 cm high, and was constructed of glass. The test section free stream velocity was adjustable over the range of 0-30 cm/sec; however, the current tests operated over only about the lower half of this range so as to ensure laminar flow conditions. The entire system contained about 3.8 m³ of water, which was recirculated and open to the atmosphere in the test section. Seeding of the water for LDV measurements was accomplished by adding about 4 gms of 1.5-2.0 μm TiO₂ particles in a water suspension every several hours. The tunnel also provided for flow visualization by means of colored dye injection.

B. Data Acquisition and Control

The data acquisition and control strategy for the flow separation tests is shown schematically in Figure 2. Separate Z-248 computers were utilized during each test; one for control of the test model operation, and another for automated data acquisition and subsequent data file storage of test results. Signal conditioning electronics associated with the LDV system consisted of a TSI Model 1990 Counter/Processor, with 12 bit 0-10 VDC D/A output proportional to inverse frequency. Sampling of the velocity signal from the Counter was accomplished by

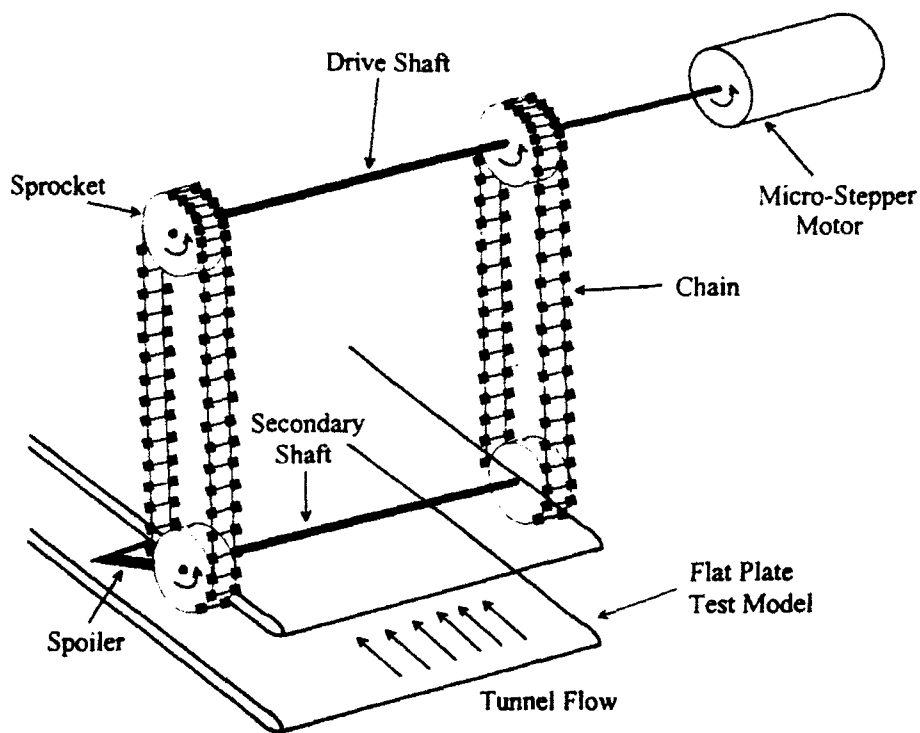


Figure 3: Schematic of Flat Plate Test Model with Spoiler Assembly

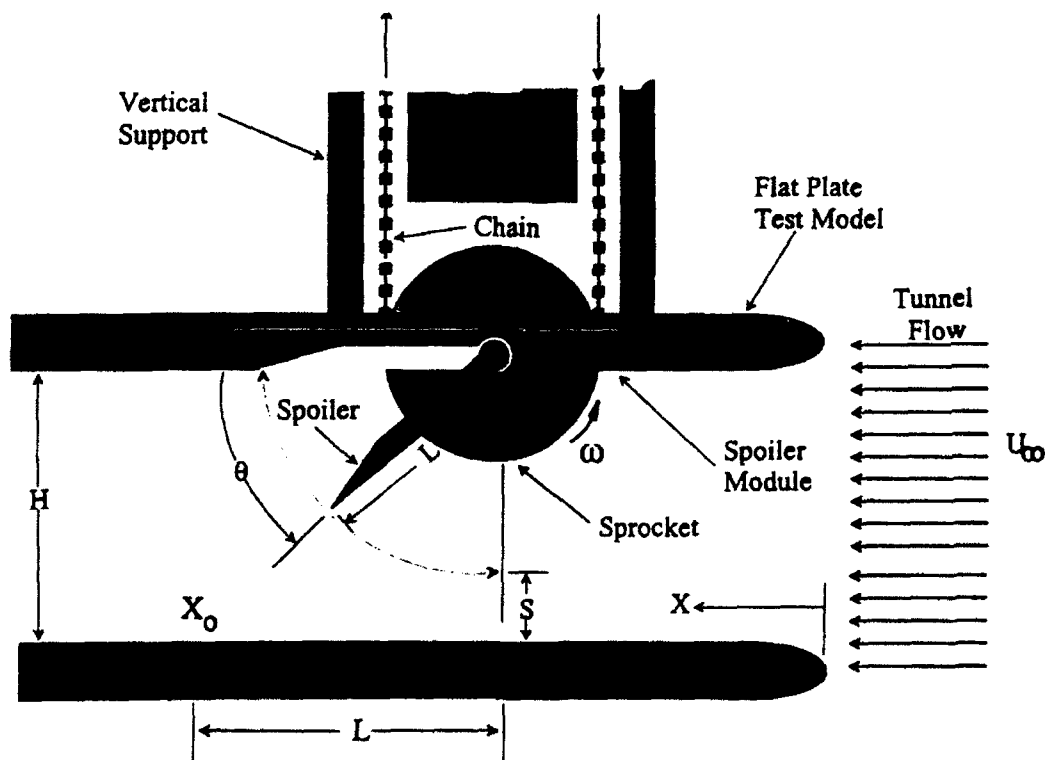


Figure 4: Detail of Test Model and Spoiler Impingement

means of a Metrabyte DAS-20 A/D board installed in the data acquisition computer.

Triggering of the data acquisition was by an internal timer on-board the DAS-20. Sampling of the velocity signal was initiated by a Gate High from one of the digital outputs available on the stepper motor indexer that controlled the spoiler flap rotation. This Gate High occurred at the beginning of the data acquisition portion of each flap rotation period (angle position 25 degrees), and lasted long enough to encompass the measurement portion of the flap cycle (a 75 degree range). The DAS-20 timer was programmed to trigger sampling at the rate of 100 samples for every 5 degrees of flap rotation; however, only the first from each 100 sample block was recorded. The high sample rate ensured minimal time error between the time of a Gate High from the motor indexer and the measurement of velocity component corresponding to that particular angular flap position.

Traversing of the LDV measuring volume was fully-automated by means of the computer-controlled 3-axis traverse. The measuring volume was moved first to a given x-position, then the vertical flow field was traversed over its entire range before proceeding to a new x-position. This procedure was an attempt to minimize the total duration of each test by reducing the transit time associated with moving the measuring volume location. Cycle-averaging of the velocity component at a given measurement location was achieved by sampling two or more flap rotation cycles at each measurement location. For the standard resolution tests, increments in the x-direction along the flat plate test model were 5.00mm; whereas in the vertical direction they were 0.50mm. For the high resolution tests, the x positions were 2.5mm apart and the y-position increment was 0.25mm.

C. Flat Plate Test Model with Rotating Spoiler Impingement

The test model used in this investigation was a smooth flat plate. It was constructed of aluminum plate approximately 100cm long, 37cm wide, 1.3cm thick and had a 16:1 elliptical leading edge. The plate surface was finished with a flat-black anodized coating. This test model was suspended below a nearly identical plate which incorporated a rotating spoiler flap on its underside, as shown in Figure 3. A computer-controlled micro-stepper motor drive was used to supply the necessary turning torque to the dual-shaft sprocket assembly of the upper plate. This upper module had been designed and tested during a previous URRP research appointment at the FJSRL (Beck, 1991). It was initially developed to investigate the formation and propagation of vortices along the underside surface of the upper plate, and to provide a dynamic flow field for testing the newly developed scanning LDV vorticity measurement system. The rotating spoiler was had a length of 5.08 cm, and was located 371.5 mm from the leading edge of the plates. It was capable of being modulated in the range of 0 to 90 degrees of arc at a cyclic frequency of slightly in excess of 1 Hz.

Figure 4 shows a close-up detail schematic of the test model (lower plate) and the spoiler flap impingement. As the spoiler was rotated from the 0 degree (flush) to the 90 degree (normal) position, an adverse pressure gradient was set up along the surface of the lower plate downstream of the spoiler. This initiated flow separation, along with subsequent vortex shedding from the boundary layer region along the upper surface of the lower plate. Two different plate spacings (89.0 mm and 76.3 mm) were investigated for the flow visualization tests; however, velocity measurements were taken only for the larger plate spacing.

IV. RESULTS

The results of flow visualization tests, and quantitative velocity component measurements for selected test conditions, are summarized below. More complete experimental results are available in an associated Master's Thesis (Berg, 1993).

An unusual number of hardware problems were encountered during this experimental investigation. These included the above mentioned laser multimodal operation, delays connected with replacement of the stepper motor indexer which failed just prior to the initial phase of quantitative tests, intermittancy of one of the counter/processor units (which contributed to the decision to modify the LDV setup to single-component operation), intermittent vertical position repeatability of the traverse (which was critical for accurate vertical positioning within the boundary layer), and malfunctions of the spoiler flap during long duration cyclic operation. The nature of the traverse repeatability problem resulted in a gradually increasing off-set of the wall zero position, depending on the number of up and down movements of the measuring volume. It was therefore necessary to limit the number of quantitative tests to a selected sample of test conditions. This was further necessitated by a flap clearance problem, which appeared after repeated cyclic operation of the rotating spoiler during flow visualization tests. The effect of the problem was minimized by not allowing the flap to fully close. Closure to within just a few degrees of the flush position significantly increased the cycle life of the test model, and a small time delay was introduced to partially compensate for the length of time that would normally have elapsed if the flap were allowed to fully close and return to the partially open position; thus preserving the same relative phasing of flap motion with respect to the cyclic flow.

A. Flow Visualization Test Results

Flow visualization tests by colored dye injection were conducted initially to aid in the interpretation of the quantitative measurements, as well as to more effectively zero in on the relevant range of the important adjustable parameters. A selected set of test conditions are shown in Figure 5 and Figure 6. Figure 5 shows the observed sequence of boundary layer separation phenomena, typical of all the tests. The boundary layer is first observed to thicken, followed by the formation of a vortex which subsequently detaches and is convected downstream.

A rather sharp, well-defined separation streamline is also readily observed from this sequence of pictures, indicating that the point of detachment of the streamline remains relatively fixed in position. The test conditions for this sequence, referred to as the base-line condition, were a nominal upstream free stream velocity of 7.6 cm/sec (0.25 ft/sec), a plate spacing of 89.0 mm, and a spoiler rotation frequency of 0.15 Hz. For these conditions, it appears that the surface streamline begins to detach from the wall at about $x-x_0 = 35\text{mm}$. Figure 6 depicts the effect of upstream velocity and spoiler rotation frequency on the boundary layer separation. Figure 6(a) depicts approximately the same relative stage of separation, but for two different levels of upstream velocity. The increased velocity level results in a separation streamline location further downstream, with approximately the same stage of separation occurring earlier in the flap rotation cycle. Figure 6(b) indicates that increasing the flap rotation frequency moves the separation point upstream, and shifts the stage of separation to later in the flap rotation cycle.

B. LDV Measurements

Both horizontal and vertical velocity component measurements were made during this investigation; however, due to space limitations, only a selected portion of the tests results will be presented here. Due to the time constraints imposed by the above mentioned hardware problems, direct vorticity measurements using the scanning LDV system were not attempted; however, a post-processing algorithm was developed which appears to preserve much of the character and many of the advantages of the scanning vorticity measurement technique. The technique will be referred to as Post-Process Scanning (PPS), and is directly related to the real-time scanning (RTS) technique developed by Beck and Humphreys (1991). The technique, depicted in Figures 7(a) and 7(b), processes the discrete spatial measurements in a manner consistent with the RTS method by scanning the velocity field in a circular path. The result is a scan-averaged velocity component, and a velocity gradient given by

$$u_a = \frac{1}{2\pi} \int_0^{2\pi} u(x,y) d\theta \quad (1)$$

$$\frac{\partial u}{\partial y} = \frac{1}{\pi r_s} \int_0^{2\pi} u(x,y) \sin \theta d\theta \quad (2)$$

The result is a very effective spatial filtering of the velocity profile, the accuracy of which is dependent on the shape of the velocity profile in the same manner as with the RTS method.

For the results presented in this report, a simple linear surface interpolation algorithm was utilized to provide local spatial velocity at positions in between the measurement grid points. In reference to Figure 7(b), this interpolation function is given by

$$u(x,y) = U_1(x) + [U_2(x) - U_1(x)] \left(\frac{y-y_1}{y_2-y_1} \right) \quad (3)$$

where

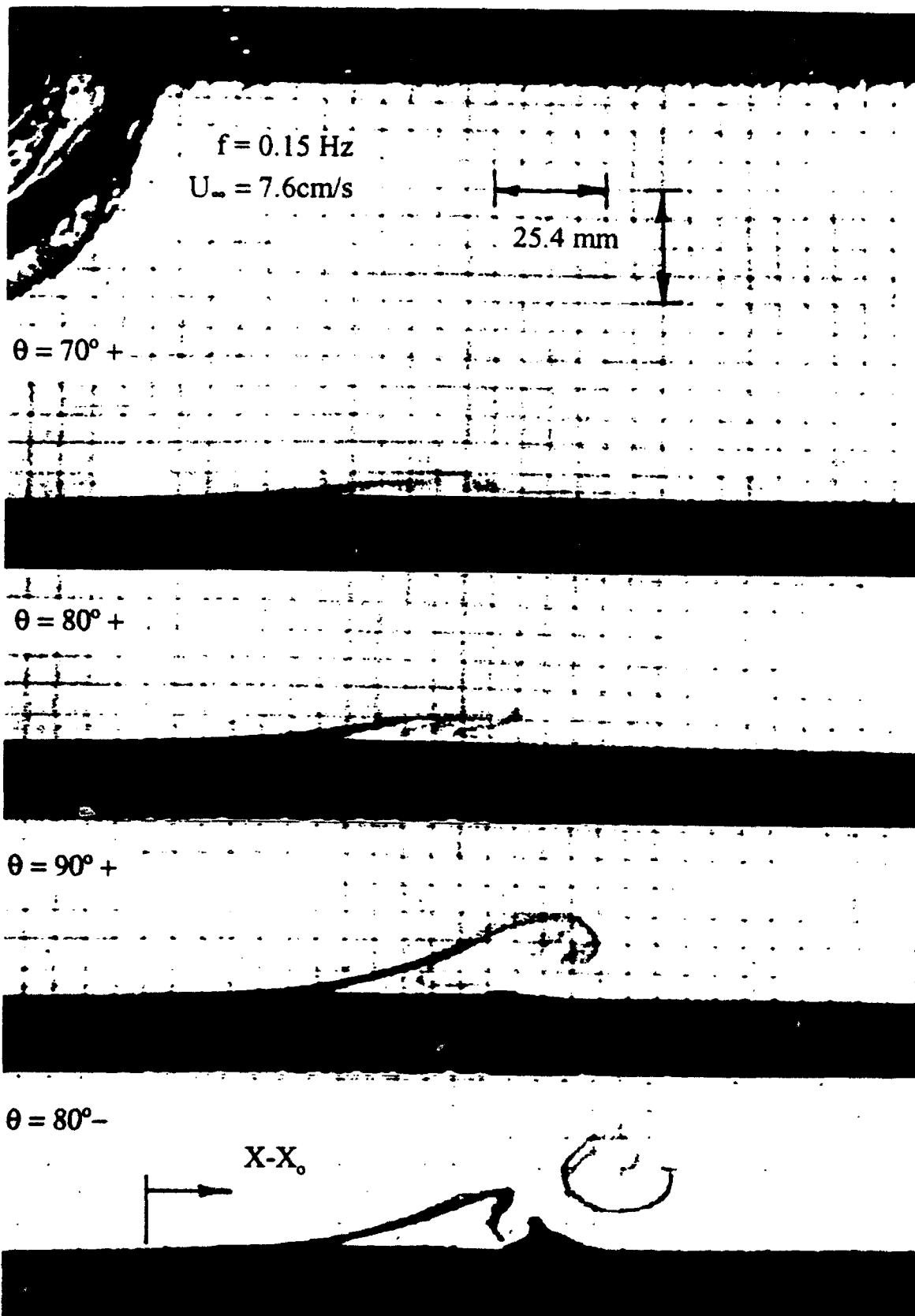
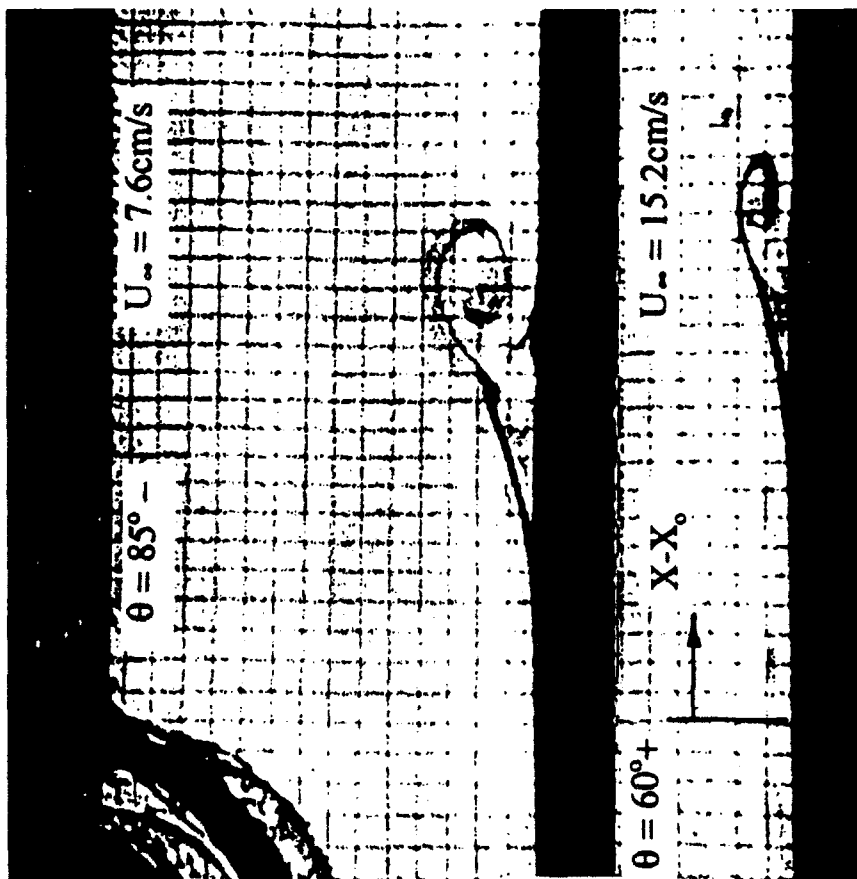


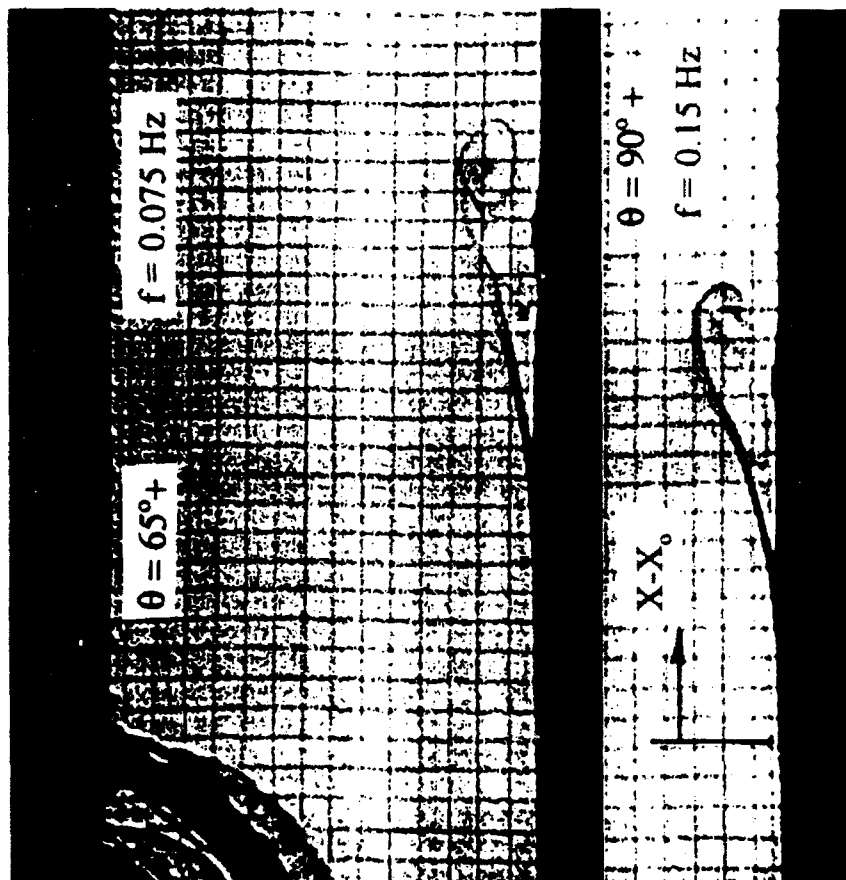
Figure 5: Flow Visualization of Boundary Layer Separation

$f = 0.15 \text{ Hz}$



(a) Effect of Free Stream Velocity

$U_\infty = 7.6 \text{ cm/s}$



(b) Effect of Flap Frequency

Figure 6: Effect of Test Parameters on Boundary Layer Separation

$$U_1(x) = u(x_1, y_1) + [u(x_2, y_2) - u(x_1, y_1)] \left(\frac{x - x_1}{x_2 - x_1} \right) \quad (4)$$

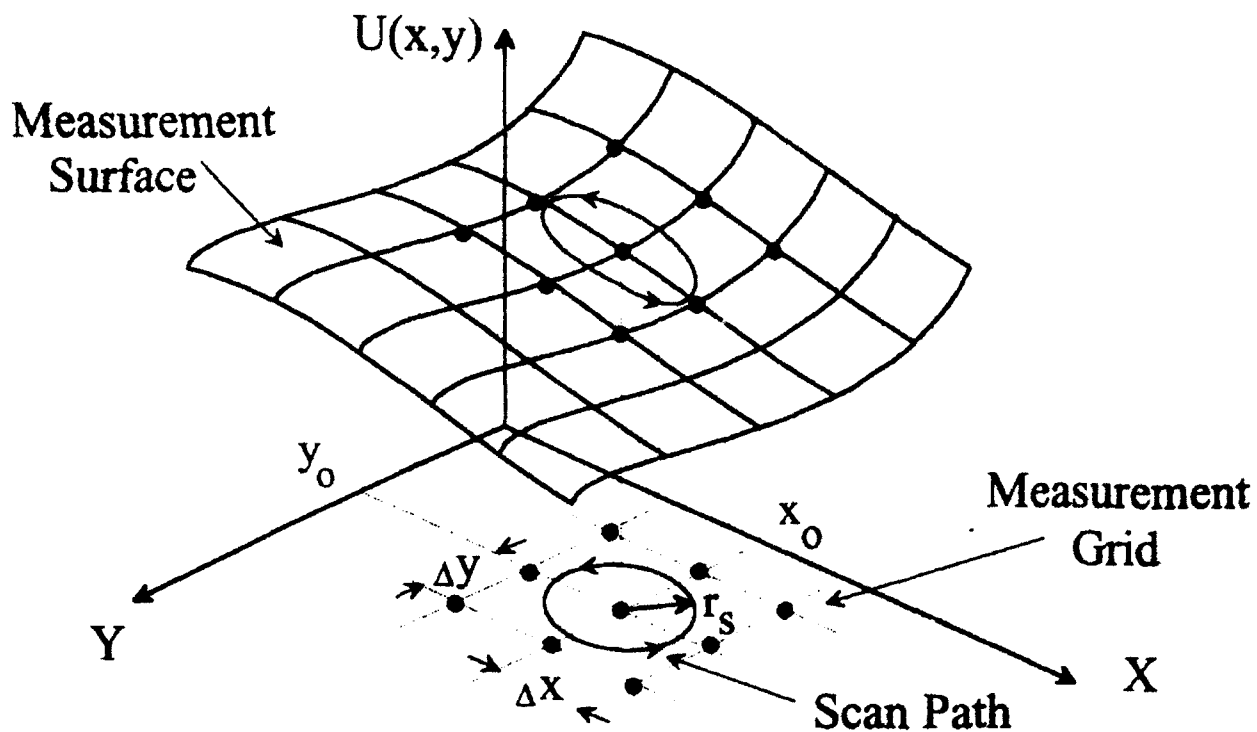
$$U_2(x) = u(x_1, y_2) + [u(x_2, y_2) - u(x_1, y_2)] \left(\frac{x - x_1}{x_2 - x_1} \right) \quad (5)$$

Figure 8(a) shows the spatial variation of the x-component of velocity at one typical angular position (instant of time), for the "base-line" test conditions. Measurements near the wall ($< .5\text{mm}$) were subject to large error due to "dazzling" of the Photomultiplier Tube by the large surface reflection. Hence, the near wall data was removed and a zero velocity condition was imposed at the wall prior to implementation of the PPS method. Figure 8(b) shows the resulting scan-averaged velocity profile processed from the data in Figure 8(a) using a scan radius of 0.75mm , which is comparable with the scan radius used in the already proven RTS method for comparable boundary layer thickness. The LDV measurements presented below depict the physical characteristics of boundary layer separation for only the base-line test conditions.

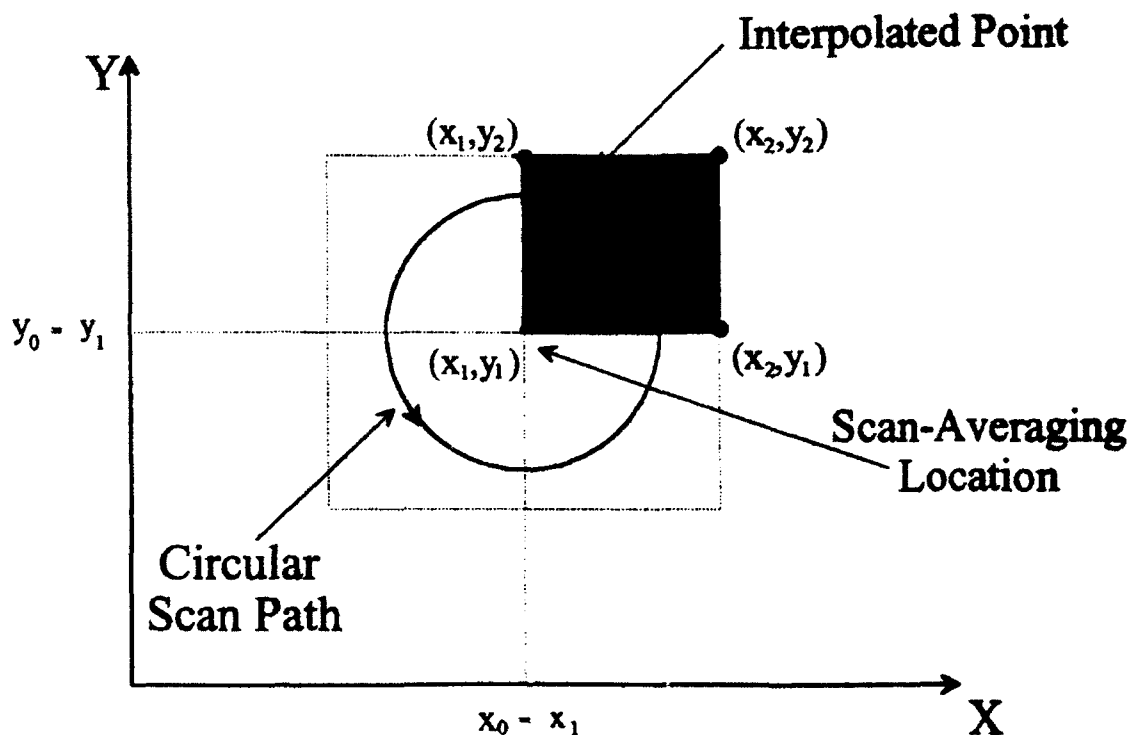
Figure 9 shows the spatial variation of the horizontal and vertical velocity components for some of the selected angle positions shown in the flow visualization pictures of Figure 5. The presence of a region of reverse flow in the x-direction is clearly evident for values of $x - x_0$ greater than about 35 mm , which is consistent with the flow visualization tests. For these test conditions it also appears that the u-component of velocity is significantly larger than the v-component, as is the gradient of u compared to v. Thus $u(x, y, t)$ provides an approximate indication of the boundary layer flow, and $\partial u / \partial y$ approximately depicts the local (and wall) shear stress as well as the vorticity distribution. More detailed analysis of results, which gives further consideration to the vertical velocity component, is given in Berg (1993).

An enlarged view for angle position 90° , which shows more detail of the near wall forward and reverse flow structure, is shown in Figure 10. Figure 11 depicts the time dependence of the vertical variation of the u-component of velocity for x-coordinate positions on each side of the position of visual streamline detachment. The depressed region of the surface plot clearly indicates the presence of reverse flows for the downstream location, but no flow reversal is evident for the upstream position.

The development of the velocity gradient (also approximately indicative of shear stress and vorticity) field is shown in Figure 12, for a sequence of four flap angle positions. The appearance of a zero wall shear boundary is clearly seen to propagate upstream, eventually remaining approximately stationary (as suggested by the approximately stationary wall streamline detachment position); thus, suggesting consistency with the basic MRS separation criteria. Finally, the profile for the last angle position shown indicates the appearance of two wall positions of zero gradient. This is consistent with the formation of a region of forward flow



(a) Post-Process Scanning (PPS) Technique



(b) Detail of Surface Interpolation

Figure 7: Scan-Averaging and Gradient Determination

next to the wall downstream of the separation streamline, as is also apparent from the enlarged near-wall view of Figure 10(b). The same sequence of events is also evident in the velocity contour map sequence shown in Figure 13. The propagation of a zero velocity level region near the wall appears to track the above zero wall gradient location described above.

V. CONCLUSIONS AND RECOMMENDATIONS

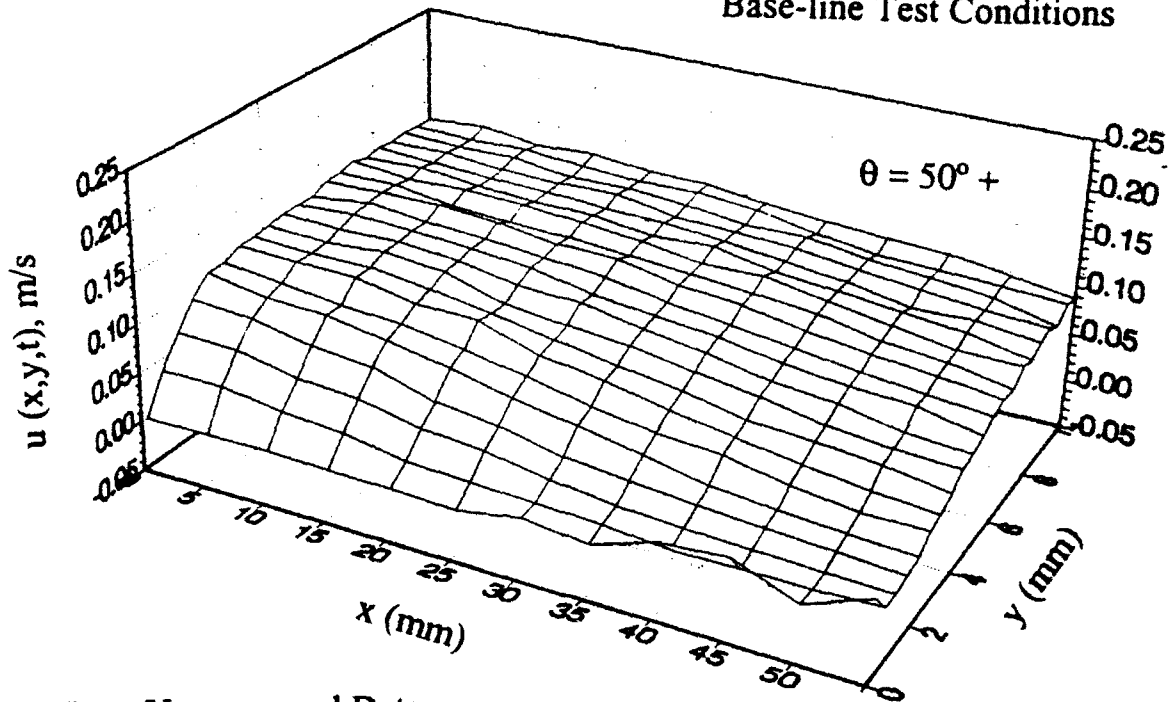
The LDV measurements in the transient spoiler-induced separating boundary layer flow clearly indicate the upstream propagation of a zero wall shear boundary. Furthermore, this appears to be consistent with the motion of the detached streamline, as seen from the flow visualization test results. In spite of the somewhat limited spatial resolution, the velocity measurements have revealed considerable detail about the flow field. Furthermore, the PPS technique has proved to be a valuable tool in both smoothing as well as determining gradients of the velocity field. The full potential of this technique is yet to be realized.

Further enhancement of the existing data utilizing the PPS method may reveal still additional detail. Intermediate interpolation of the velocity data is also possible with PPS, as is repeated application of the method to determine (and/or smooth) higher order spatial (or even time) derivatives. However, more detailed spatial and time resolution should still be investigated; particularly, in the near-wall region. Counter-based signal processing has very limited capability very near the wall due to "dazzling" of the photomultiplier and the low signal to noise ratio. Therefore, it is recommended that modern autocorrelation processors be utilized to study the near wall phenomena. The excellent repeatability of the spoiler-induced boundary layer separation also makes it an excellent flow field for such a more detailed investigation.

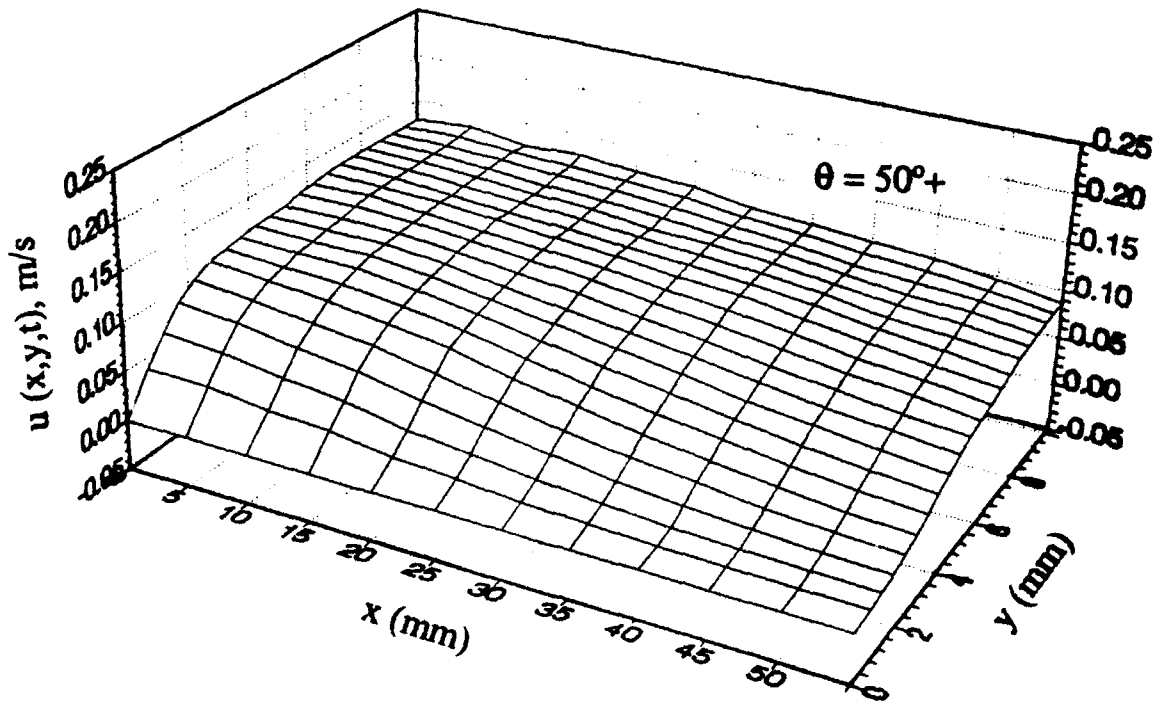
VI. REFERENCES

- Didden, N. and Chih-ming Ho, May 1985, "Unsteady separation in a boundary layer produced by an impinging jet," *J. Fluid Mech.*, Vol. 160, pp. 235-256.
- Beck, B.T. and W.W. Humphreys, Aug. 1991, "A Scanning LDV Vorticity Measurement System," presented at the 4th International Conference on Laser Anemometry, Cleveland, Ohio, and published in the Proceedings, Vol. 1, pp. 105-114.
- Reynolds, W.C. and L.W. Carr, 1985, "Review of Unsteady, Driven Separated Flows," *AIAA Paper 85-0527*.
- Gad-el-Hak, M. and D. Bushnell, January 1991, "Status and Outlook of Flow Separation Control," *AIAA Paper AIAA-91-0037*, 29th Aerospace Sciences Meeting, Reno, Nevada.
- Lovato, J.A., 1992, "Active Control of the Separation Region on a Two-Dimensional Airfoil," Frank J. Seiler Research Laboratory, Report# FJSRL-TR-92-0001.
- Sears, W.R., 1956, "Some Recent Developments in Airfoil Theory," *J. Aeronaut. Sci.*, Vol. 23, pp. 490-499.

Base-line Test Conditions



(a) Raw Unprocessed Data



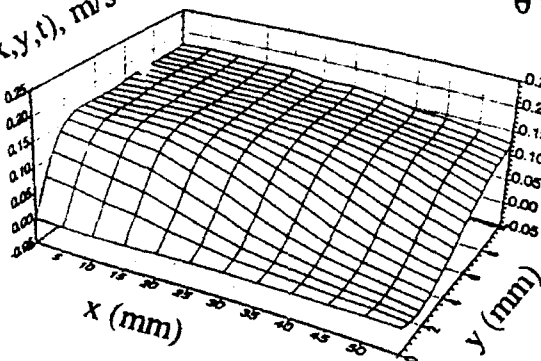
(b) Scan-Averaged Data

Figure 8: Effect of Scan-Averaging on $u(x,y)$

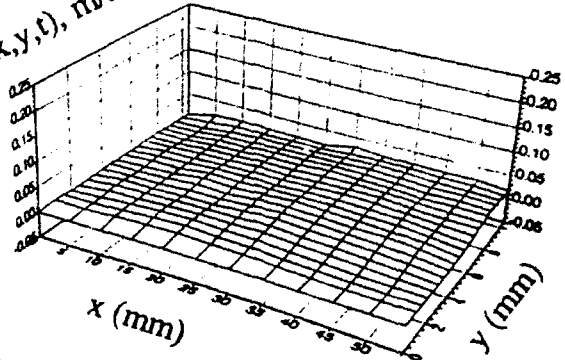
Base-line Test Conditions

$\theta = 65^\circ+$

$u(x,y,t)$, m/s

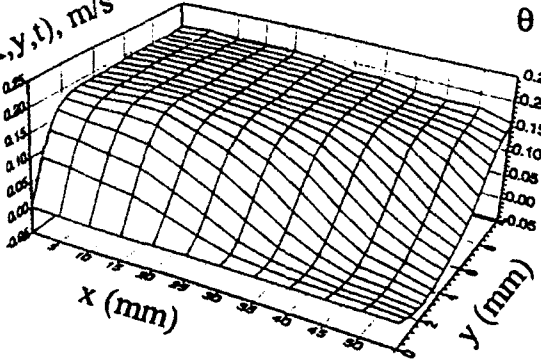


$v(x,y,t)$, m/s

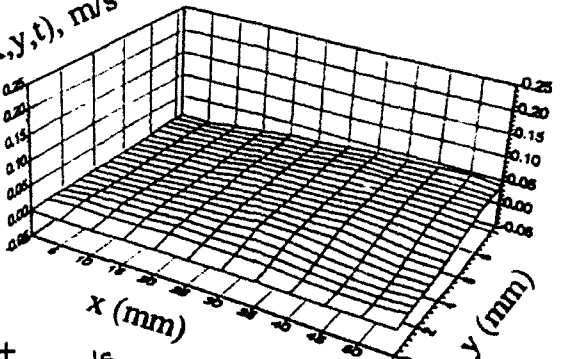


$\theta = 75^\circ+$

$u(x,y,t)$, m/s

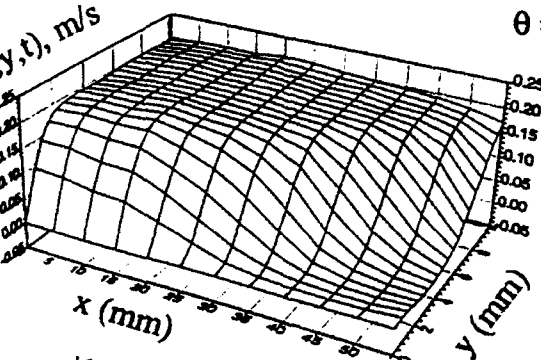


$v(x,y,t)$, m/s

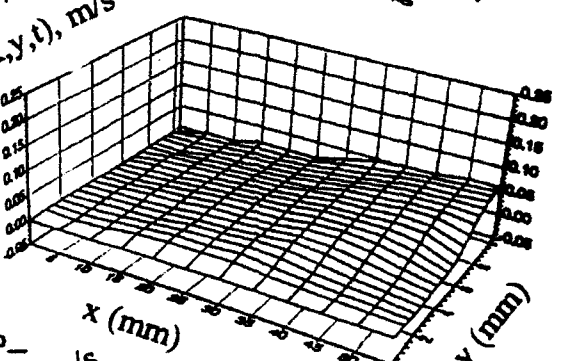


$\theta = 85^\circ+$

$u(x,y,t)$, m/s

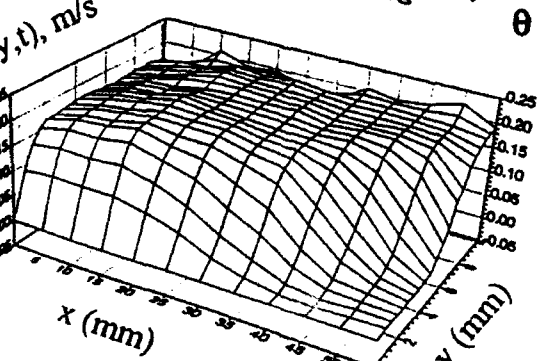


$v(x,y,t)$, m/s

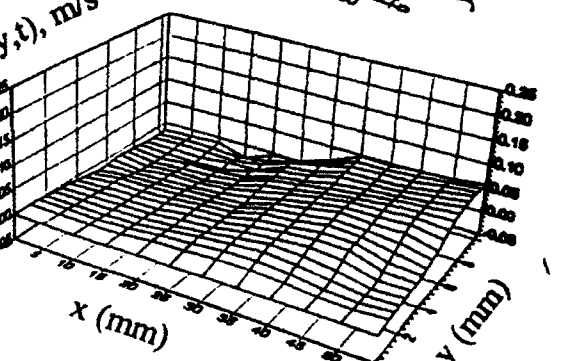


$\theta = 85^\circ-$

$u(x,y,t)$, m/s



$v(x,y,t)$, m/s



(a) Horizontal Component

(b) Vertical Component

Figure 9: Spatial Distributions of Horizontal and Vertical Velocity Components

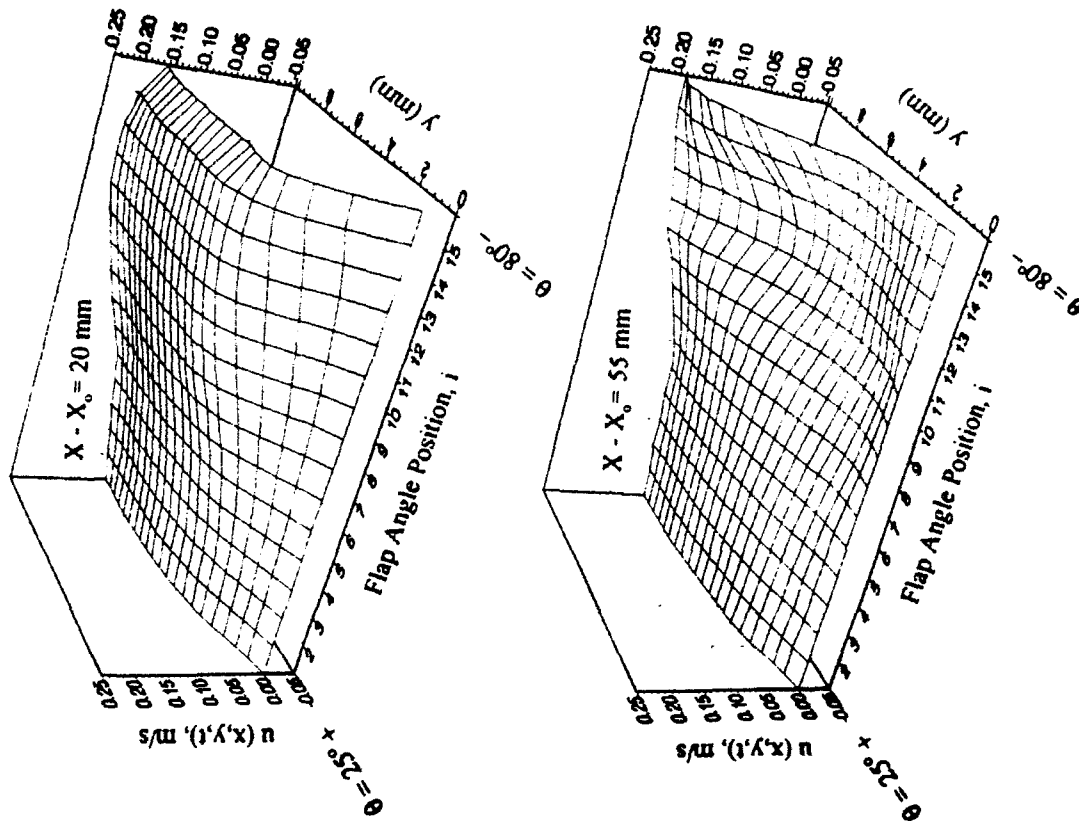


Figure 11: Transient Response Upstream/Downstream of Streamline Detachment

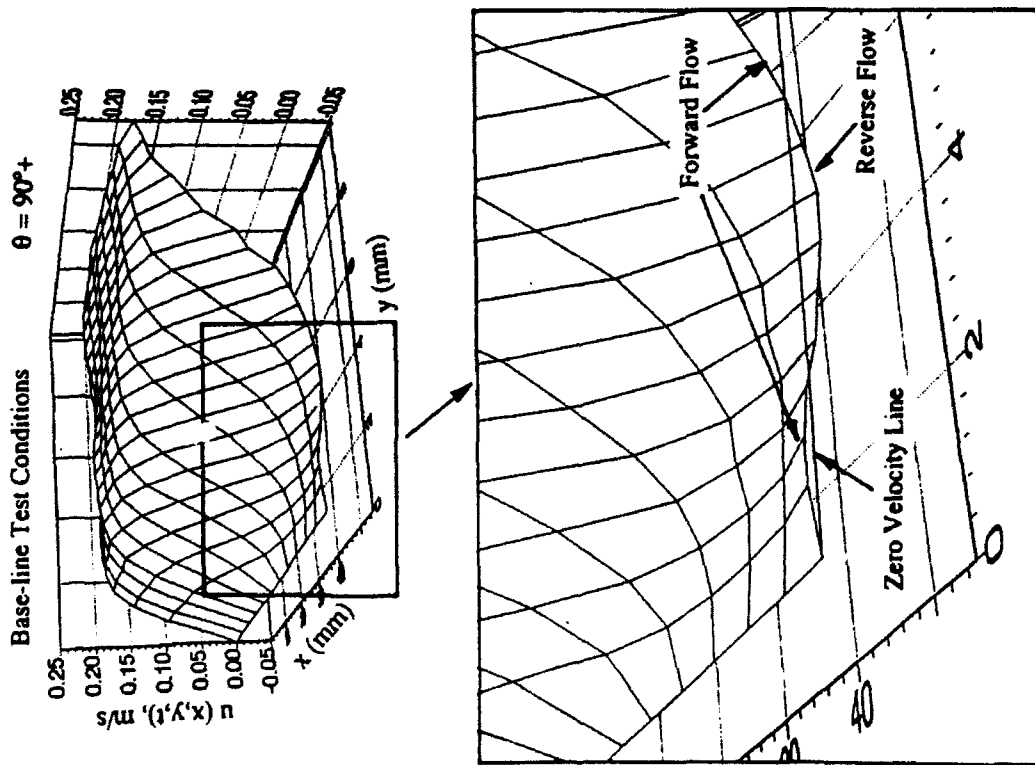


Figure 10: Detail of Near-Wall Flow Structure

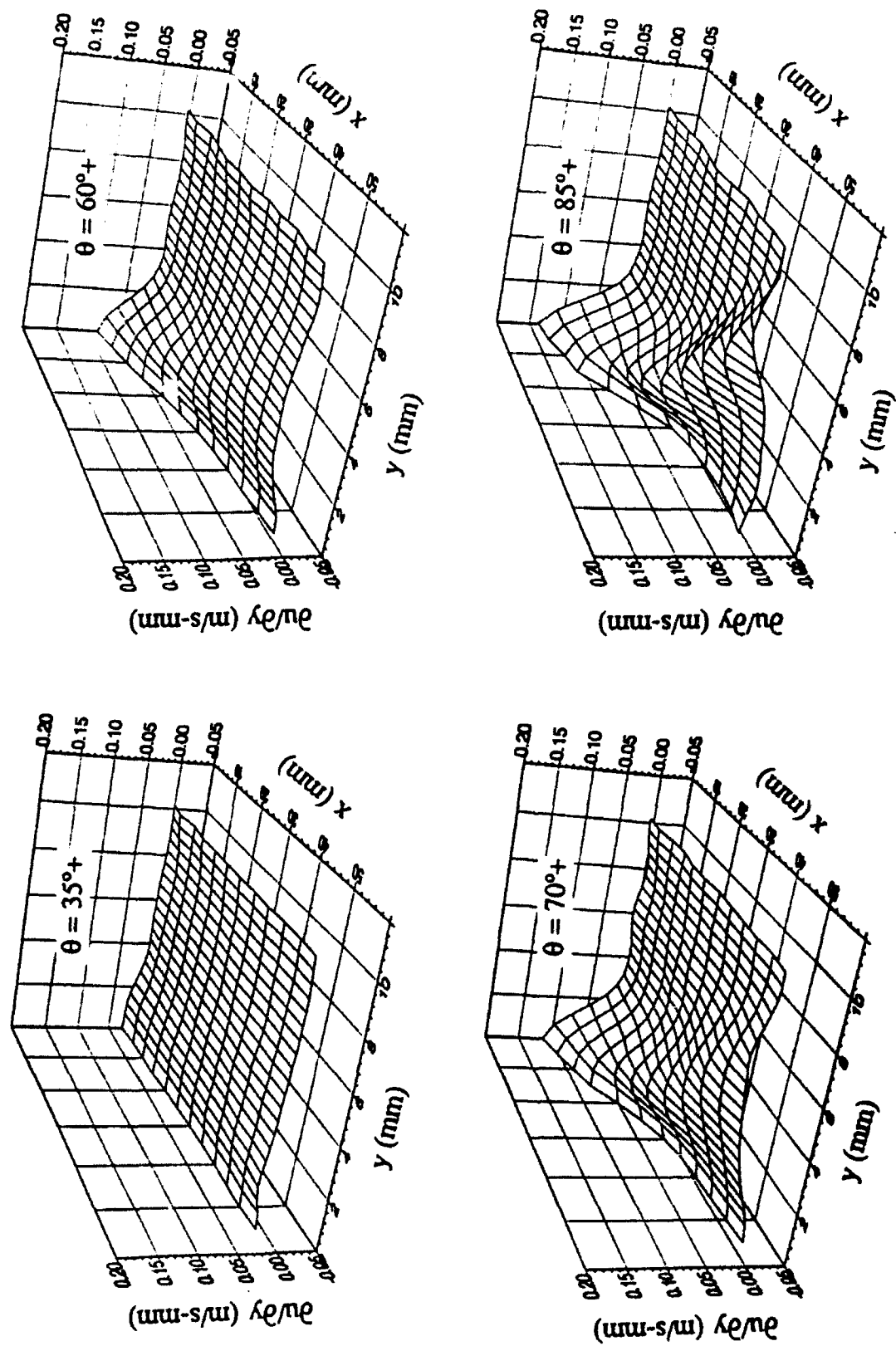


Figure 12: Spatial Distribution of Velocity Gradient

Berg, Paul K., 1993, "Velocity and Vorticity Measurements in Unsteady Separating Boundary Layer Flows," Master's Thesis (in preparation).

VII. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the cooperation of the personnel and staff of the FJSRL, and those of the Aeronautics Laboratory at the USAF Academy, who made this research possible. Special thanks are also given to the Summer Research focal point Capt. Scott Schreck for his valuable input and suggestions, to Mr. Bobby Hatfield for machining of the test model and modification of the existing hardware, to Sgt. Steve Ramsey for his assistance with hardware communications, to Dr. Julie Lovato for valuable consultations, and to Sgt. Rick Durmon for his continuous assistance with the water tunnel facility.

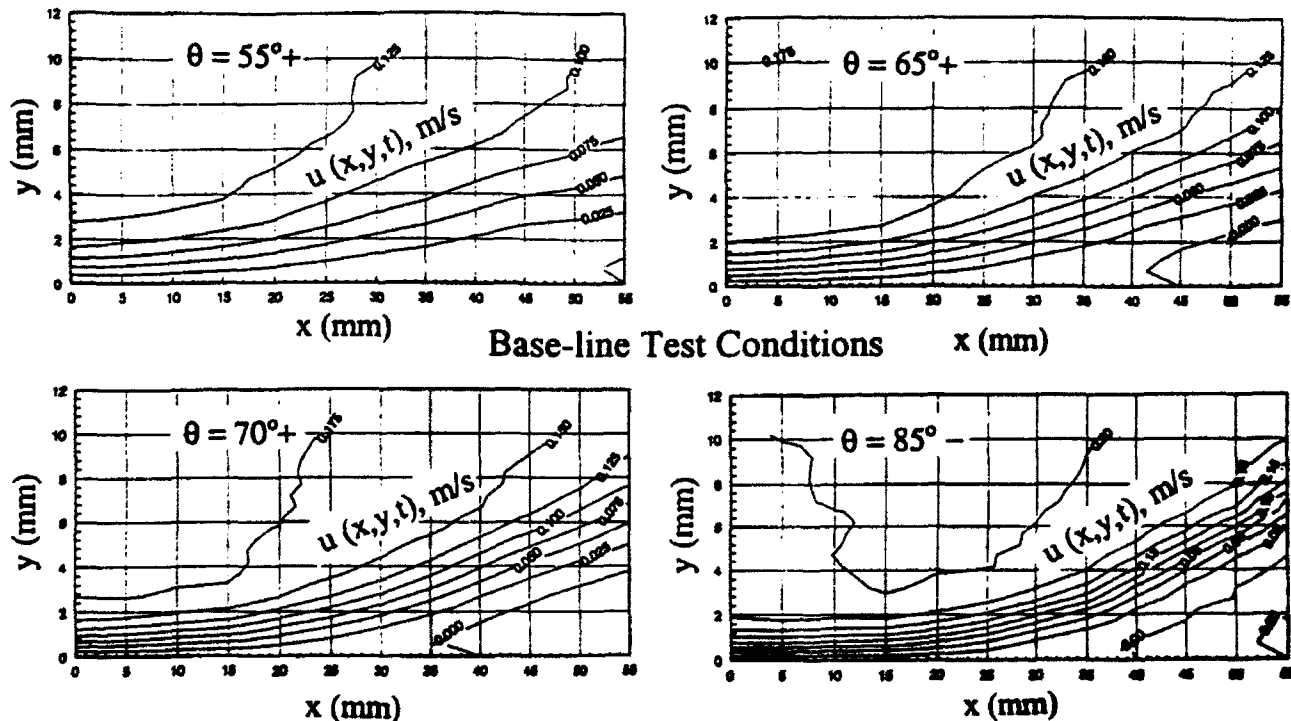


Figure 13: Horizontal Velocity Component Contour Maps

**NMR RELAXATION STUDIES OF MICRODYNAMICS IN
CHLOROALUMINATE MELTS**

**William R. Carper
Professor
Department of Chemistry**

**Wichita State University
Wichita, Kansas 67208**

**Final Report for:
Summer Research Program
Frank J. Seiler Research Laboratory,
USAF Academy, CO 80840-6528**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

August 1992

NMR Relaxation Studies of Microdynamics in
Chloroaluminate Melts

William R. Carper
Professor
Department of Chemistry
Wichita State University

Abstract

The microdynamics of EtAlCl_2 containing melts are examined by ^{13}C NMR relaxation methods as a function of melt composition and temperature. Application of the Dual Spin Probe (DSP) method to these systems reveals interaction between (1) the MEI^+ methyl group, (2) the terminal CH_3 of the MEI^+ ethyl group, and various EtAlCl_2 containing species. Unlike MEICl-AlCl_3 room temperature melts, there is no indication of interaction between the MEI^+ ring $\text{CH}'\text{s}$ and EtAlCl_2 .

INTRODUCTION

Room temperature molten salts consisting of mixtures of AlCl_3 and 1-ethyl-3-methylimidazolium chloride (MEICl), are of interest as aprotic solvents for studying a wide range of both organic and inorganic compounds [1-7]. These chloroaluminate molten salts possess considerable potential as battery electrolytes and various types of electrochemical agents [8-10].

The composition of a chloroaluminate melt has a considerable effect on its physical properties. The variations in physical properties of the melt are due to a combination of factors including ion-ion interactions [4], and Lewis acid-base properties. Chloroaluminate melts with AlCl_3 present in excess (mole fraction, N , of AlCl_3 , > 0.5) are termed acidic with AlCl_4^- and Al_2Cl_7^- the predominant anions.

The use of NMR relaxation methods provides useful information about the dynamics and structure of various chemical systems and chloroaluminate systems in particular. In a previous work [11], ^{13}C NMR relaxation measurements were used to investigate the motion and interactions of the MEI^+ cation. The results indicate that AlCl_4^- in a $\text{Na}^+_{0.22}\text{MEI}^+_{0.78}\text{AlCl}_4^-$ melt forms a complex by interacting with the C-2, C-4 and C-5 hydrogens on the MEI^+ ring. This investigation was followed by studies [12,13] in which the Dual Spin Probe method [14] supported the existence of $\text{MEI}(\text{AlCl}_4)_n^{(n-1)+}$ complexes in neutral ($\text{AlCl}_3 = \text{MEICl}$) and NaCl -buffered melts. ^{27}Al , ^{23}Na and ^{13}C NMR relaxation results confirmed the presence of the chloroaluminate- MEI^+ complexes and yielded ^{27}Al and ^{23}Na liquid state quadrupole coupling constants [12,13].

Application of the Dual Spin Probe (DSP) relaxation method typically requires knowledge of ^{13}C dipolar relaxation rates which are defined by (1), the basic equation in which the ^{13}C nucleus is relaxed by ^1H [15]:

$$R_1^{\text{dd}} = N_H (\hbar \gamma_C \gamma_H)^2 r_{\text{CH}}^{-6} \tau_{\text{eff}} \quad (1)$$

where R_1^{dd} ($= 1/T_1^{\text{dd}}$) is the dipolar relaxation rate, N_H is the number of

hydrogens attached directly to the carbon atom, γ_C and γ_H are gyromagnetic ratios and $r_{CH} = 1.09 \text{ \AA}$. τ_{eff} is the effective correlation time and varies exponentially with temperature. Equation (1) is operative while under the "extreme narrowing condition" ($\omega\tau_{eff} \ll 1$) which is usually applicable for small molecules including the chloroaluminate melts[11].

R_1^{dd} is obtained by measuring T_1 , the Nuclear Overhauser Enhancement factor, η ($\eta_{max} = \gamma_H/2\gamma_C$) and using eqn (2) [16]:

$$R_1^{dd} = \eta R_1 / 1.988 \quad (2)$$

The other part of the DSP method requires knowledge of quadrupolar relaxation rates for nuclei such as ^{27}Al and ^{23}Na . If there is a distortion from tetrahedral or cubic symmetry, nuclei such as ^{27}Al and ^{23}Na will be under the influence of an electric field gradient which produces the quadrupole interaction. The quadrupolar relaxation rate in the "extreme narrowing region" is given by(3) [15,17]:

$$R_1 = [3\pi^2(2I+3)/10I^2(2I-1)][1+(z^2/3)][e^2Qq/h]^2\tau_c \quad (3)$$

where $I = 3/2$ for ^{23}Na and $5/2$ for ^{27}Al , eQ is the nuclear quadrupole moment, eq is the maximum component of the electric field gradient tensor, and z is the asymmetry parameter of the electric field gradient tensor($z = 0$ for AlCl_3).

The quadrupole coupling constant, QCC, is given by:

$$\text{QCC} = [e^2Qq/h] \quad (4)$$

The DSP method has been applied to chloroaluminate melts[12,13] and has provided evidence that the ring hydrogens of MEI^+ interact with the tetrachloroaluminate anion. The existence of these complexes has been supported by linear plots of ^{13}C dipolar relaxation rates(R_1^{dd}) vs. quadrupolar ^{27}Al

relaxation rates(R_1) that pass through the origin as predicted by equation (5):

$$R_1^{dd}(^{13}\text{C})/N_H(\gamma_C\gamma_H)^2r_{CH}^{-6} = R_1(^{27}\text{Al})/\alpha\chi^2 \quad (5)$$

where $\alpha = [3\pi^2/10][(2I + 3)/I^2(2I - 1)][1 + (z^2/3)]$, and $QCC = \chi$.

During this summer research program, the DSP method was applied to melts containing MEICl , AlCl_3 , and EtAlCl_2 . The inclusion of EtAlCl_2 provided a "baseline" as there is a covalent bond between the ethyl group and aluminum in EtAlCl_2 . The existence of covalent bonding(or complexation) between quadrupolar and dipolar nuclei in a molecule results in a linear plot of eqn. (5) that passes through the origin. In the MEICl-EtAlCl_2 melts reported herein, we observed a linear plot of eqn (5) that passed through the origin when applied to the terminal CH_3 carbon in EtAlCl_2 and one of the peaks in the ^{27}Al NMR of the melts.

EXPERIMENTAL

Materials

The 1-ethyl-3-methylimidazolium chloride (MEICl) and chloro-aluminate molten salts were prepared as described previously [1]. Ethylaluminum dichloride (EtAlCl_2) was obtained from Aldrich. All materials were stored under anhydrous helium gas atmosphere in a dry box. All molten salt preparations and manipulations were performed in the dry box. Samples were loaded into 5 mm sample tubes, capped in the dry box, removed, and sealed immediately with a torch.

NMR Measurements

^{13}C NMR spin-lattice relaxation times were recorded this summer on a Varian Gemini-300 spectrometer at 75.43 MHz and ^{27}Al NMR spin-lattice relaxation times were recorded previously on a Varian XL-300 spectrometer operating at 78.15 MHz. Temperature measurements were calibrated against methanol or ethylene glycol and

are accurate to within 0.5°C. Pulse widths(90°) were typically 8.6 (75.43 MHz) and 7.6(78.15 MHz) μ s. Longitudinal relaxation times were measured by the the inversion-recovery method (180° - τ - 90° -T) with $T > 10T_1$. At least 12 delay times(τ) were used and the results fitted to a three parameter exponential. ^{13}C NOE measurements were made using the gated decoupler method[18]. It is likely that the error in the NOE measurements is in the 5-10% range[18].

RESULTS AND DISCUSSION

The ability of both AlCl_3 and EtAlCl_2 to form $\text{C}_{2\text{H}}$ dimers[19,20] led us to examine the ^{27}Al spectra of: (1) neat EtAlCl_2 , (2) MEICl-EtAlCl_2 and (3) ternary melts ($N = \text{AlCl}_3/\text{MEICl}/\text{EtAlCl}_2$)[21]. The neat EtAlCl_2 ^{27}Al NMR spectrum contains two peaks [21]. Peak 1 is a broad downfield peak that dominates the spectrum. The second peak (upfield) overlaps peak 1 and is only a fraction of peak 1 in total peak area. Peak 2 collapses into peak 1 as the temperature is lowered from 60 to 25°C. These two aluminum sites are consistent with the extent of monomer-dimer formation in liquid EtAlCl_2 [21].

The MEICl-EtAlCl_2 ($N = 0.5/0.5$) melt ^{27}Al NMR spectrum also has two peaks. In this case, peak 1(downfield) is very broad while peak 2 is very sharp, and has a low peak area. Peak 2 increases slightly in area and peak 1 broadens as the temperature is lowered from 70 to 0°C. We have previously[21] made the tentative assignments of EtAlCl_2 for peak 1(downfield) and $\text{Et}_2\text{Al}_2\text{Cl}_4$ for peak 2.

In this study, we first apply the DSP method to the CH_3 carbon in EtAlCl_2 and ^{27}Al NMR peaks 1 and 2 from several melt combinations and neat EtAlCl_2 . Fig. 1 contains the results for ^{27}Al peak 1(downfield) and Fig. 2 contains the results for ^{27}Al peak 2. The fact that both plots are linear and pass through the origin, indicate that: (1) the DSP method is appropriate for these systems and (2) the species associated with each peak contains EtAlCl_2 . Furthermore, the slopes of these lines can be used to calculate the relative quadrupole coupling constants for the EtAlCl_2 -containing species in solution.

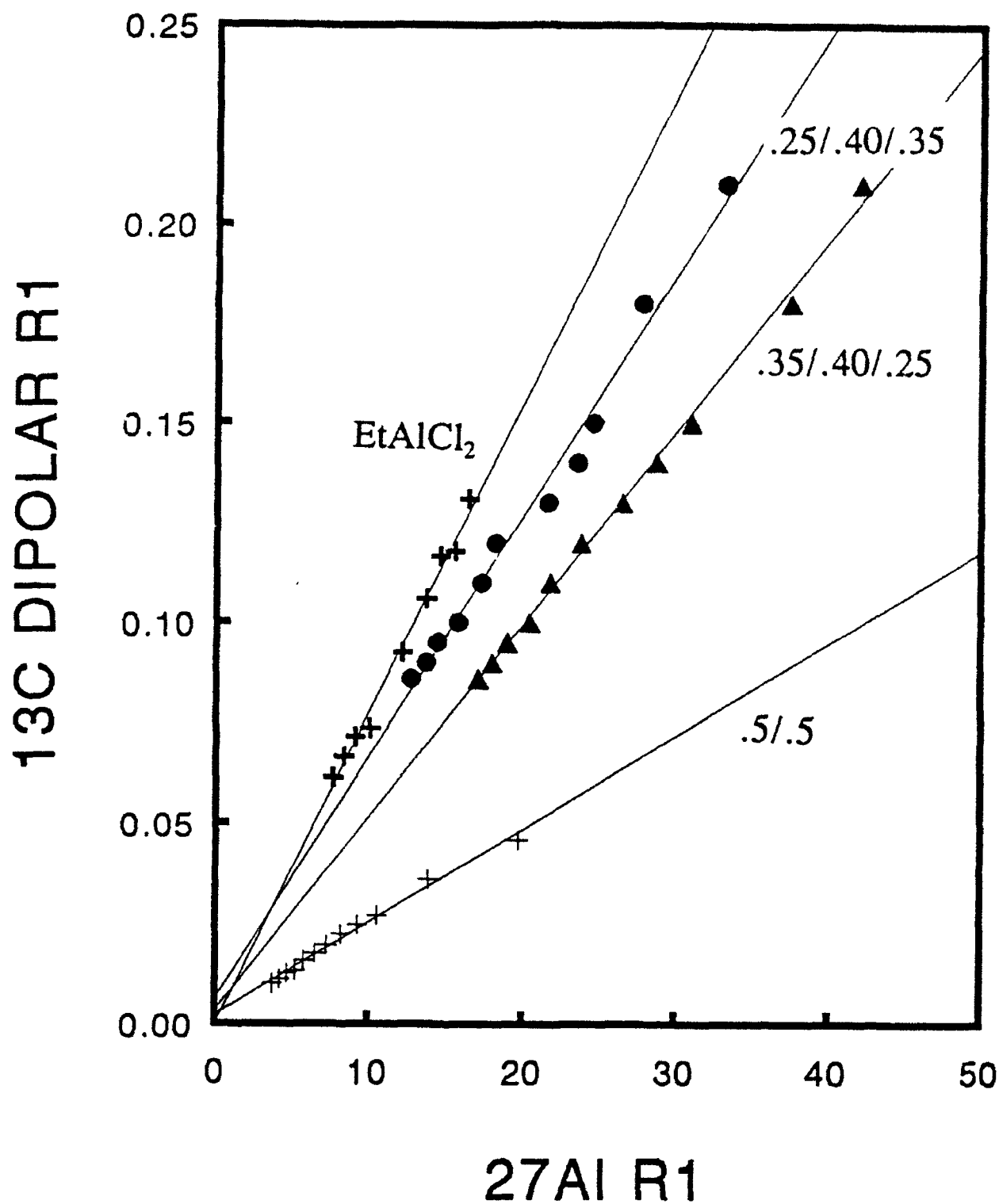


Fig. 1. ¹³C Dipolar R1's vs ²⁷Al R1's (25 to 70°C) for Al peak 1 (127-131 ppm from Al(H₂O)₆³⁺).

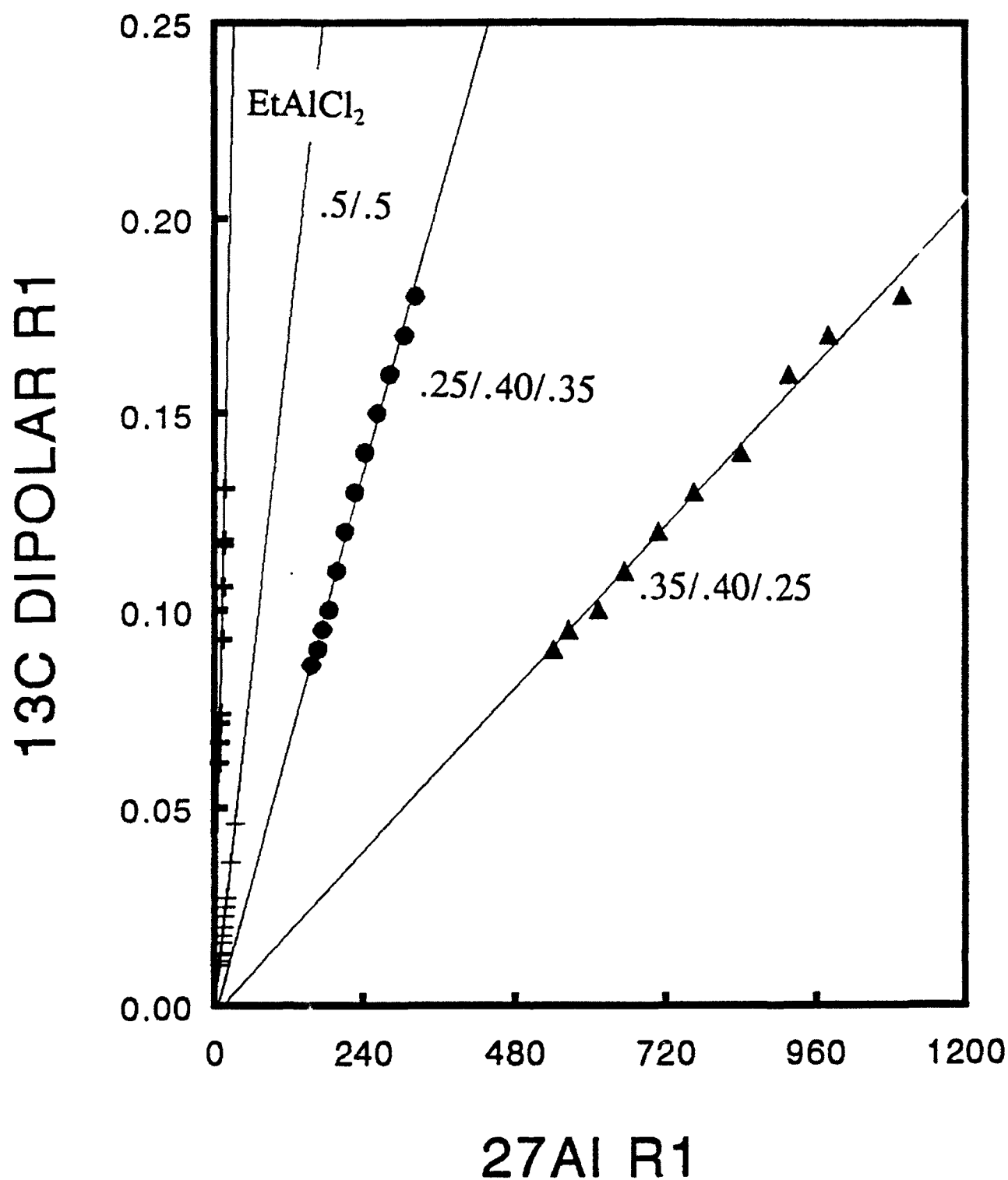


Fig. 2. ¹³C Dipolar R1's vs ²⁷Al R1's (25 to 70°C) for Al peak 2 (102.5-103.0 ppm from Al(H₂O)₆³⁺).

The QCC values obtained from Fig. 1 (Al peak 1) are 171, 119, 106 and 93 MHz for the (.5/.5), (.35/.40/.25), (.25/.40/.35) melts and neat EtAlCl₂, respectively. The QCC values obtained from Fig. 2 (Al peak 2) are 6.9, 20, 11 and 93 MHz for the (.5/.5), (.35/.40/.25), (.25/.40/.35) melts and neat EtAlCl₂ (repeated).

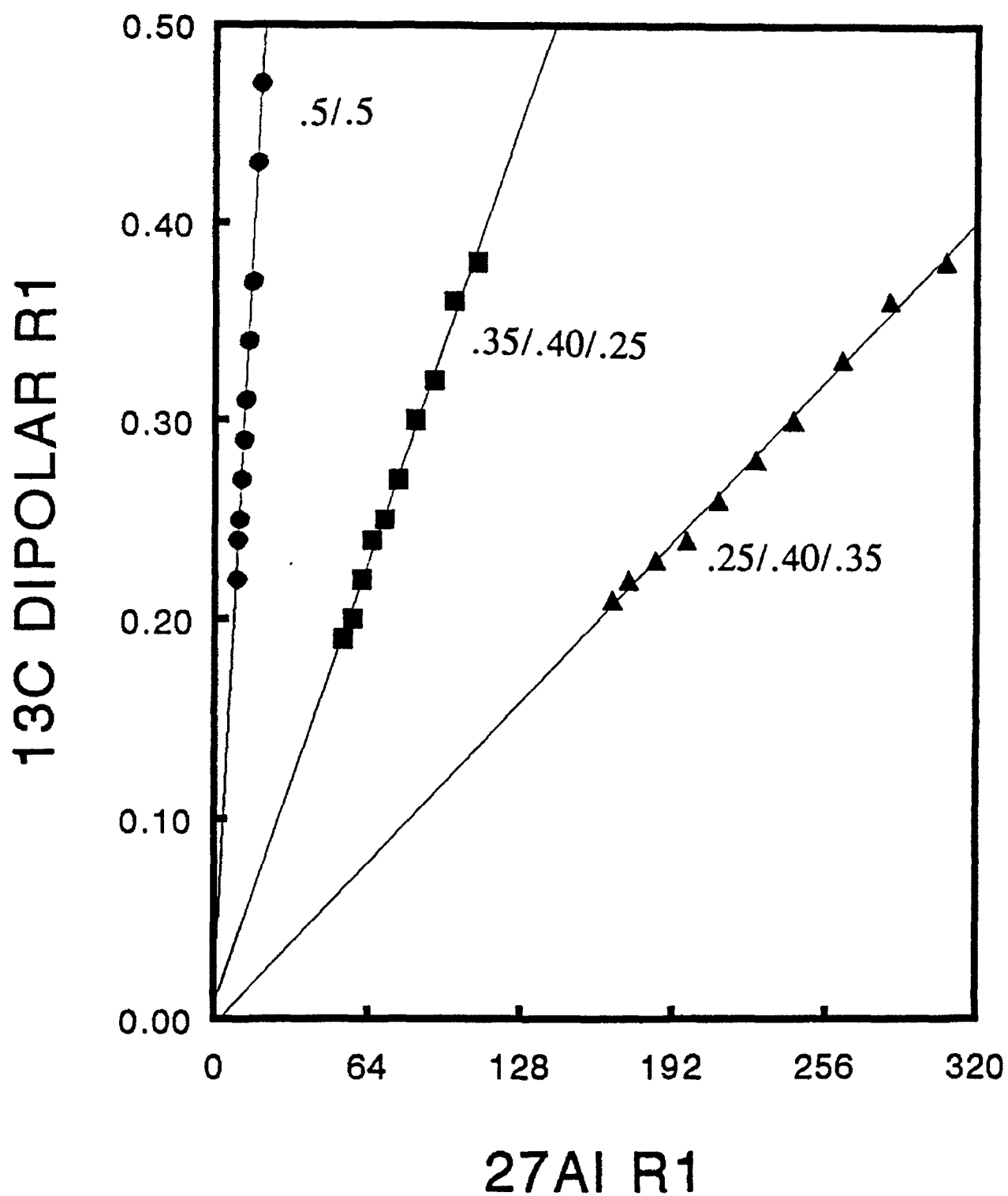
Results of the Dual Spin Probe method (eqn. [5]) applied to the (.5/.5), (.35/.40/.25) and (.25/.40/.35) melts indicate interactions between the Al-containing species in peak 2 (102.5-103.0 ppm relative to Al(H₂O)₆³⁺) and both the NCH₃ and ethyl terminal CH₃ groups of MEI⁺. Fig. 3 contains the plots for the NCH₃ group in each melt and Fig. 4 contains data for the terminal CH₃ on the MEI ethyl group.

The QCC's obtained from the slopes in Fig. 3 (MEI NCH₃) are 1.7, 2.3 and 4.4 MHz for the (.5/.5), (.35/.40/.25) and (.25/.40/.35) melts. The QCC's from Fig. 4 (terminal CH₃ on the MEI ethyl group) are 1.6, 6.9 and 1.3 MHz for the (.5/.5), (.35/.40/.25) and (.25/.40/.35) melts.

Finally, there is no correlation between the ring hydrogen dipolar R1's and any of the ²⁷Al peak R1's. This result is directly opposite to that found in MEICl-AlCl₃ systems [11,12].

CONCLUSIONS

Application of the DSP method to these mixed melt systems indicates a lack of complexation between the ring hydrogens of MEI⁺ and any of these aluminum containing species. These and previous results [21] suggest that the formation of various charged dimers containing EtAlCl₂ takes precedence over the formation of complexes between EtAlCl₂ and the MEI⁺ ring hydrogens. However, there is evidence of interaction between the various Al-containing species and the CH₃ groups (NCH₃ and terminal CH₃ in the ethyl group) of MEI⁺ in these melts.



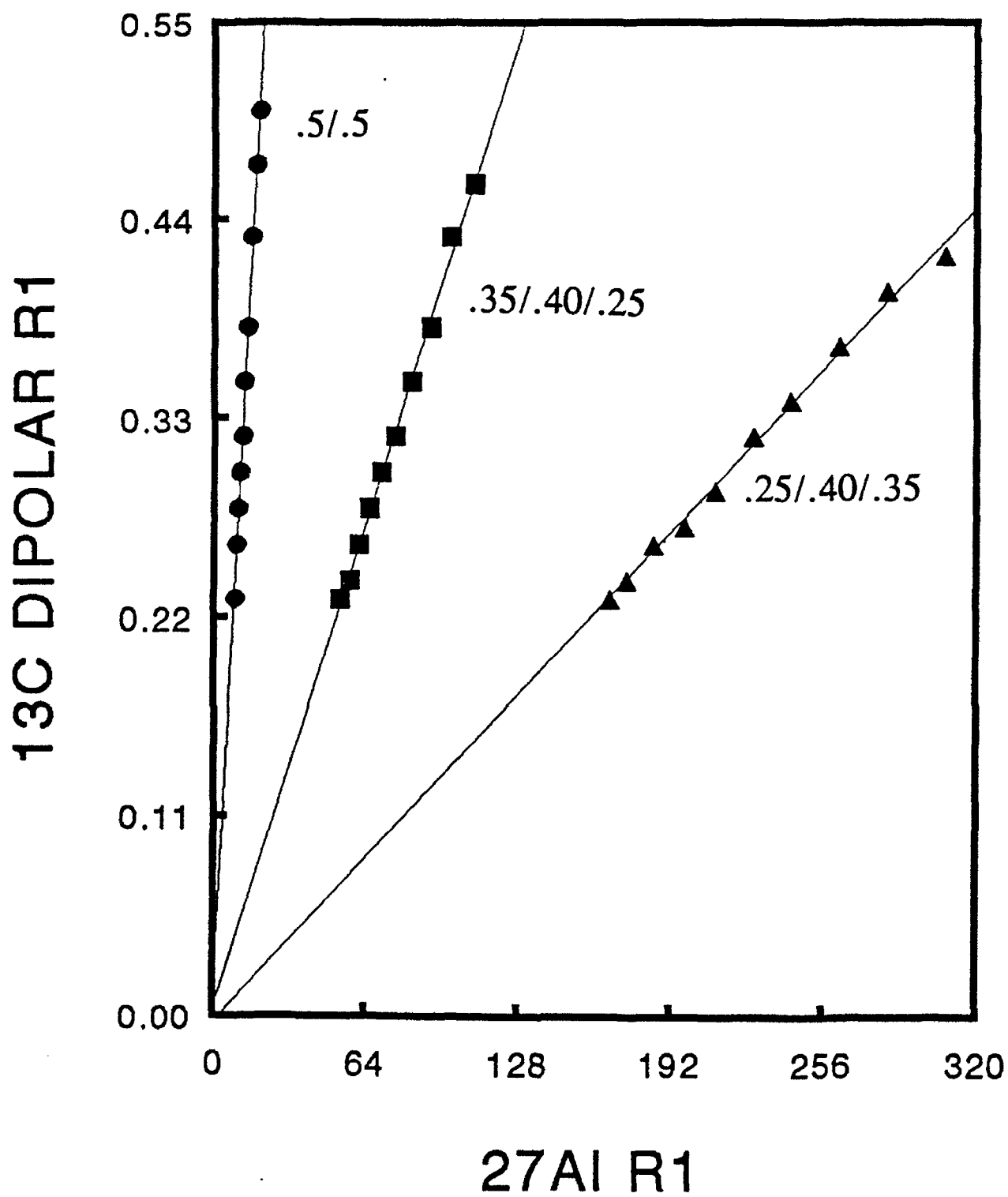


Fig. 4. ^{13}C Dipolar R1's for ethyl CH_3 carbon vs ^{27}Al R1's (25 - 70°C) for Al peak 2.

REFERENCES

- [1] J. S. Wilkes, J. A. Levisky, R. A. Wilson and C. L. Hussey, *Inorg. Chem.*, **21** 1263 (1982).
- [2] J. S. Wilkes, J. S. Frye and G. F. Reynolds, *Inorg. Chem.*, **22** (1983) 3870.
- [3] A. A. Fannin, L. A. King, J. A. Levisky and J. S. Wilkes, *J. Phys. Chem.*, **88** (1984) 2609.
- [4] A. A. Fannin, D. A. Floreani, L. A. King, J. S. Landers, B. J. Piersma, D. J. Stech, R. L. Vaughn, J. S. Wilkes and J. L. Williams, *J. Phys. Chem.*, **88** (1984) 2614.
- [5] K. M. Dieter, C. J. Dymek, N. E. Heimer, J. W. Rovang and J. S. Wilkes, *J. Amer. Chem. Soc.*, **110** (1988) 2722.
- [6] C. J. Dymek and J. J. P. Stewart, *Inorg. Chem.*, **28** (1989) 1472.
- [7] J. A. Boon, J. A. Levisky, J. L. Pflug and J. S. Wilkes, *J. Org. Chem.*, **51** (1986) 480.
- [8] C. J. Dymek, J. L. Williams, D. J. Groeger and J. J. Auborn, *J. Electrochem. Soc.*, **131** (1989) 2887.
- [9] C. J. Dymek and L. A. King, *J. Electrochem. Soc.*, **132** (1985) 1375.
- [10] C. L. Hussey, T. B. Scheffler, J. S. Wilkes and A. A. Fannin, *J. Electrochem. Soc.*, **133** (1986) 1389.
- [11] W. R. Carper, J. L. Pflug, A. M. Elias and J. S. Wilkes, *J. Phys. Chem.* **96** (1992) 3828.
- [12] W. R. Carper, J. L. Pflug and J. S. Wilkes, *Inorganica Chimica Acta* **193** (1992) 201.
- [13] W. R. Carper, J. L. Pflug and J. S. Wilkes, *Inorganica Chimica Acta* (in press).
- [14] J. J. Dechter and U. Henriksson, *J. Magn. Res.*, **48** (1982) 503.
- [15] A. Abragam, "Principles of Nuclear Magnetism", Oxford University Press, Oxford (1961).
- [16] K. F. Kuhlmann and D. M. Grant, *J. Amer. Chem. Soc.*, **90** (1968) 7355.

- [17] B. Lindman and S. Forsen, in "NMR Basic Principles and Progress," P. Diehl, E. Fluck and R. Kosfeld, Editors, Vol. 12, p. 22, Springer-Verlag, New York (1976).
- [18] D. Neuhaus and M. Williamson, "The Nuclear Overhauser Effect in Structural and Conformational Analysis", VCH Publishers, New York (1989).
- [19] J. Weidlein, *J. Organomet. Chem.*, 17 (1969) 213.
- [20] B. Gilbert, Y. Chauvin and I. Guibard, *Vib. Spectros.*, 1 (1991) 299.
- [21] W. R. Carper, C. E. Keller, P. A. Shaw, M. P. and J. S. Wilkes, in "Eighth International Symposium on Molten Salts", Electrochem. Soc., New York (in press).

**CHARACTERIZATION OF MATERIALS
FOR NON-LINEAR OPTICAL THIN FILMS**

**Thomas M. Christensen
Assistant Professor
Department of Physics**

**University of Colorado at Colorado Springs
P.O. Box 7150
Colorado Springs, CO 80933**

**Final Report for:
AFOSR Summer Research Program
Frank J. Seiler Research Laboratory**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington D.C.**

September 1992

CHARACTERIZATION OF MATERIALS FOR NON-LINEAR OPTICAL THIN FILMS

Thomas M. Christensen
Assistant Professor
Department of Physics
University of Colorado at Colorado Springs

Abstract

Recent progress in non-linear optical materials has opened up tremendous opportunities for the development of inexpensive, non-linear optical devices. For this to occur, we need to better understand the origins of the non-linear optical properties of these materials. This is facilitated by being able to make thin films of these materials with desirable properties. We have explored the chemical composition, defect concentration and non-linear optical properties of thin films of GeO_x and SiO_2 doped with GeO_x produced by laser ablation and sputter deposition. The GeO_x films produced by laser ablation in a vacuum are laterally non-uniform with Ge enrichment at the center while those produced in an oxygen environment are more uniform and more like GeO_2 . A photoluminescence experiment was set up, but the small sample volume available for analysis in thin films probably produced too small a luminescence to detect. A method for periodically poling films externally to enhance second harmonic generation is presented along with some calculations of field penetration into the films.

CHARACTERIZATION OF MATERIALS FOR NON-LINEAR OPTICAL THIN FILMS

Thomas M. Christensen

INTRODUCTION

Second harmonic generation is an inherently non-linear optical phenomenon in which light incident on a material with a particular frequency, ω , leaves the material with a mix of frequencies, ω and 2ω . It has long been known that this phenomenon is not allowed in materials which are centro-symmetric. It was a surprise, therefore, in late 1985 when second harmonic generation was observed [1] from amorphous SiO_2 fibers doped with GeO_2 . These glass fibers were believed to be centro-symmetric, yet a second harmonic signal was observed to grow in over the course of several hours. It was later observed [2] that by "seeding" the fibers with the 2ω light, the second harmonic generation could be enhanced. Since the index of refraction in a material depends on the frequency of the light, the ω and 2ω light should propagate at different velocities through the fiber and thus be out of phase with one another. The strong observed second harmonic signal indicated that phase matching was occurring. This could be accomplished by creating a spatial periodicity in the material. In the seven years since the first observations, a great deal of effort has gone into trying to explain the source and formation of the non-linear properties in fibers. Many models have been proposed, but no consensus has yet been reached on the fundamental physics behind these effects.

More recently, similar effects have been observed [3] in thin films of the same material. The observation in thin films opens up new possibilities in characterizing the materials in order to understand the underlying physics. In this report, we examine these materials in terms of chemical composition (including spatial non-uniformities), defect concentration, and non-linear optical properties. The report is divided into four main sections discussing each of the four research projects which were attempted.

CHEMICAL COMPOSITION AND NON-UNIFORMITY

Auger Electron Spectroscopy

The chemical composition of films prepared by both sputtering and laser ablation was examined with Auger electron spectroscopy (AES) which uses a beam of electrons to excite

atoms in a material. The deexcitation process can occur by either emitting electrons (AES) or light (luminescence - described in the next section). The emitted electrons come off with energies which are characteristic of the elements from which they came. Slight shifts in these energies can be used to determine the chemical state of the atom (metallic vs. oxide). AES is a surface sensitive technique with information coming primarily from the top 30 Å of the material.

We obtained spectra using a standard Perkin-Elmer single pass cylindrical mirror analyzer. Ge and Si emit Auger electrons at both low (40-100 eV) and high (1000-1600 eV) energies. Charging problems associated with a beam of electrons directed onto an insulating surface forced us to keep the incident beam energy low (1000 eV). As a result the high energy transitions were not excited. The Auger transitions used here were GeO_x (45 eV), SiO_2 (76 eV), C (271 eV), and O (510 eV). All peaks were shifted to reference them to C at 271 eV for those spectra where charging had caused anomolous peak shifts.

Typical spectra of a SiO_2 film nominally doped with 5 % GeO_2 are shown in Figure 1. This film was produced by sputter deposition at Phillips Laboratory and was approximately 2 microns thick. The concentration of Ge in the film surface can be estimated from peak height measurements and standard sensitivity factors [4]. We estimate that this film is 6 ± 3 % GeO_x in good agreement with the nominal value. The difference in energy between GeO_2 and GeO is too small to accurately say anything about the specific oxidation state of the material, however it is definitely oxidized and not metallic.

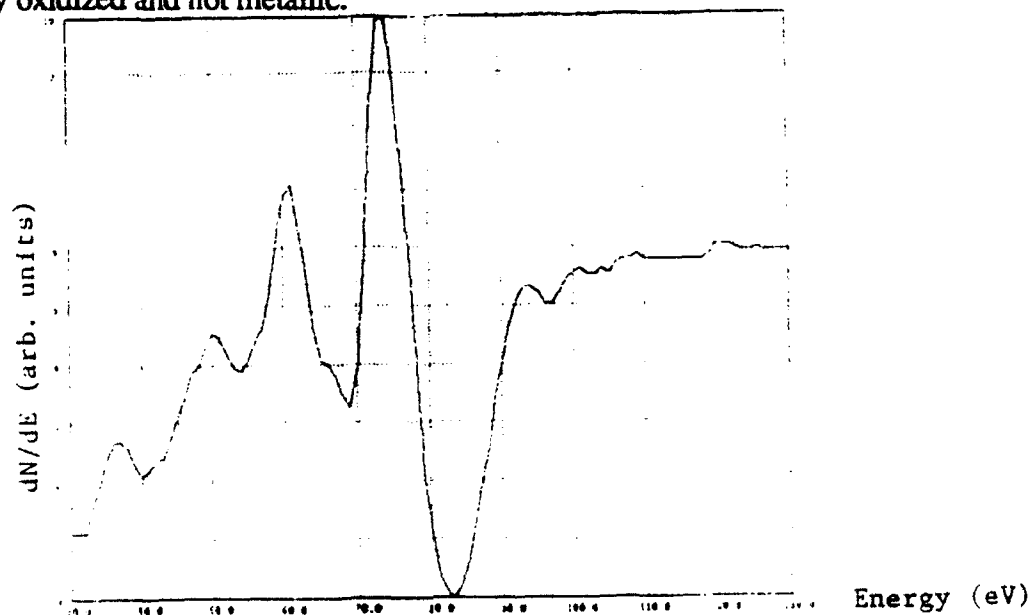


Figure 1. AES spectrum of a GeO_2 doped SiO_2 film.

Films are presently being made at Seiler Laboratory using laser ablation. This process is known to have a deposition plume in which different species can be spatially separated. This can lead to a spatially non-uniform film. We have examined the composition and lateral uniformity of films produced by laser ablation in both vacuum and oxygen environments. The results are presented in Figure 2. The peak to peak height ratios are used rather than atomic concentration ratios because of uncertainty in the sensitivity factors. Films grown in vacuum are observed to have a higher Ge/O ratio near the center of the film with decreasing amounts of Ge as you move toward the edges. Films grown in an oxygen environment (100 mTorr) are much more uniform and have an overall lower Ge/O ratio (as would be expected for GeO_2 rather than GeO).

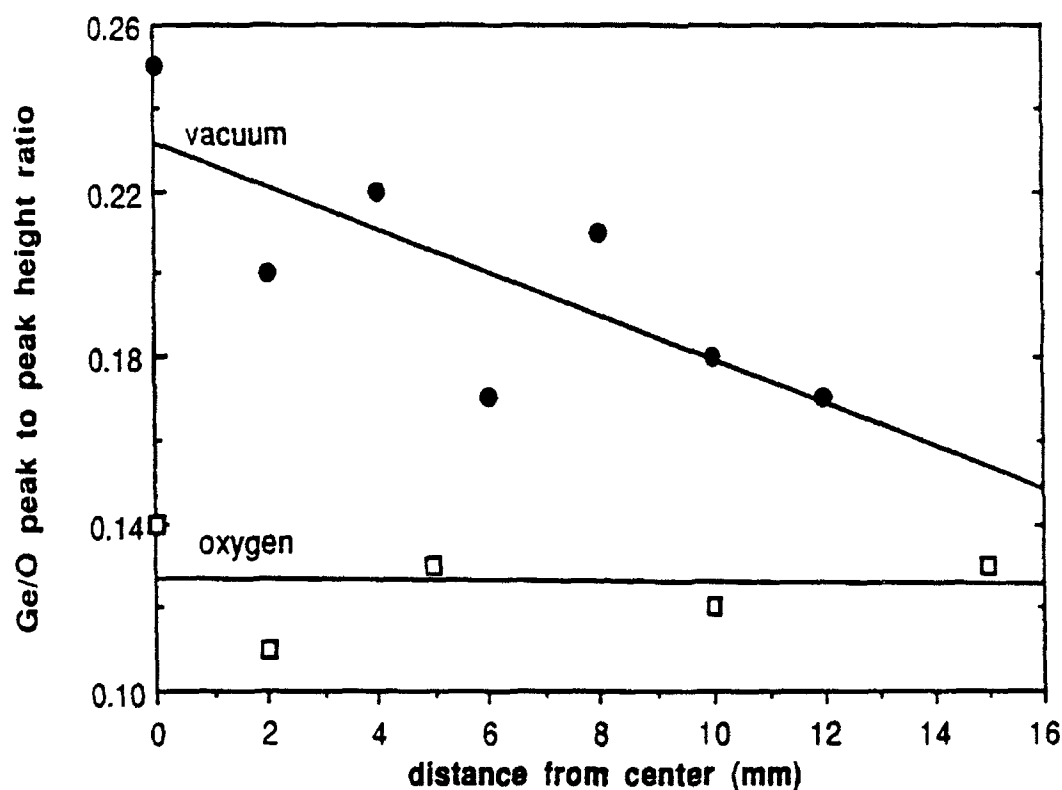


Figure 2. Ge/O AES peak height ratio vs. lateral distance from center for film produced by laser ablation in vacuum (•) or in 100 mTorr of oxygen (□).

We also examined several films made using laser ablation from mixed GeO_2 and SiO_2 targets. Several trends were apparent. All of the films which we examined showed a Ge

enrichment at the edges of the film. All of the films produced in 100 mTorr of O₂ look basically identical, even when the target composition was changed from 80% SiO₂ to 20% SiO₂. This suggests that the surface of the film may not be representative of the bulk since IR measurements on these samples did look different. The one film which we examined that was made in vacuum had a much higher Ge/Si ratio.

We also observed that films made in an oxygen environment also had more carbon in the film. This seemed to be incorporated into the film and not just on the surface, suggesting that the chamber or gas lines might be somewhat dirty. We examined a couple of samples made at Sieler Lab using a sputter deposition system there. These samples were supposed to be pure GeO and pure GeO₂. We found that both had small levels of Si contamination.

DETECTING DEFECTS - LUMINESCENCE

(Part of this portion of the study was done in collaboration with Dr. Ralph Hill, U.S. Air Force Reserves.)

Defects in fibers and films are believed to be involved in establishing the second harmonic generating properties of these materials. We sought to be able to examine defects so that we could understand the fundamental processes involved in the establishment of non-linear optical properties and so that we could characterize the materials being produced. Two methods are typically used for looking at defects in ceramic and glass materials. One of these is electron spin resonance. This technique will be pursued by other Seiler personnel. We examined using luminescence to detect defects in thin films.

Luminescence is another deexcitation process (competing with the Auger process) in which an electron which has been excited into the conduction band falls back down into the valence band with the excess energy released in the form of a photon. If the excitation is accomplished by light, as in our studies, the process is known as photoluminescence. Traditionally, photoluminescence, involved using a light source with an energy greater than the band gap of the material being studied. If high power lasers are available, electrons can be excited by less efficient, multi-photon processes. By examining the spectrum of light emitted from the material, we seek to observe features associated with various defects in the material. The existence of defect levels within the band gap provides alternate paths for deexcitation. At room temperature these features will be rather broad and weak. With bulk materials, cooling to liquid helium temperatures is common to obtain sharp, strong signals. This procedure presents

difficulties, however, in working with thin films because of the possible differences in thermal expansion coefficients between the substrate and film. If the two contract differently, stress may be induced into the film which will alter the luminescence being studied.

We tried both single photon (direct) and multi-photon processes for exciting the luminescence in the thin films. We will first discuss the multi-photon excitation and then the direct.

Several lasers were used to attempt multiphoton excitation. Continuous lasers ranging from low power He-Cd and Kr lasers up to an Argon ion laser ($\lambda = 488$ or 514 nm) was used as the source of excitation for most of the studies. The Ar lasere was operated at powers ranging from 10 to 1600 mWatts, which yielded an effective power at the sample of approximately $1/8$ of the operating power due to various losses in the optics. The light from the laser was chopped to allow use of a lock-in amplifier to detect faint signals. A 12.5x microscope objective lens was used to focus light from the sample onto the entrance slit ($250 \mu\text{m}$) of the monochromator. A photodetector was placed immediately outside the exit slit ($100 \mu\text{m}$) of the monochromator. Several photodetectors were used with increasing sensitivity to detect smaller signals. The present configuration had ample sensitivity to detect the rather large background signal of stray light from the monochromator. This signal was substantially reduced if the laser was blocked. A typical configuration is shown in Figure 3.

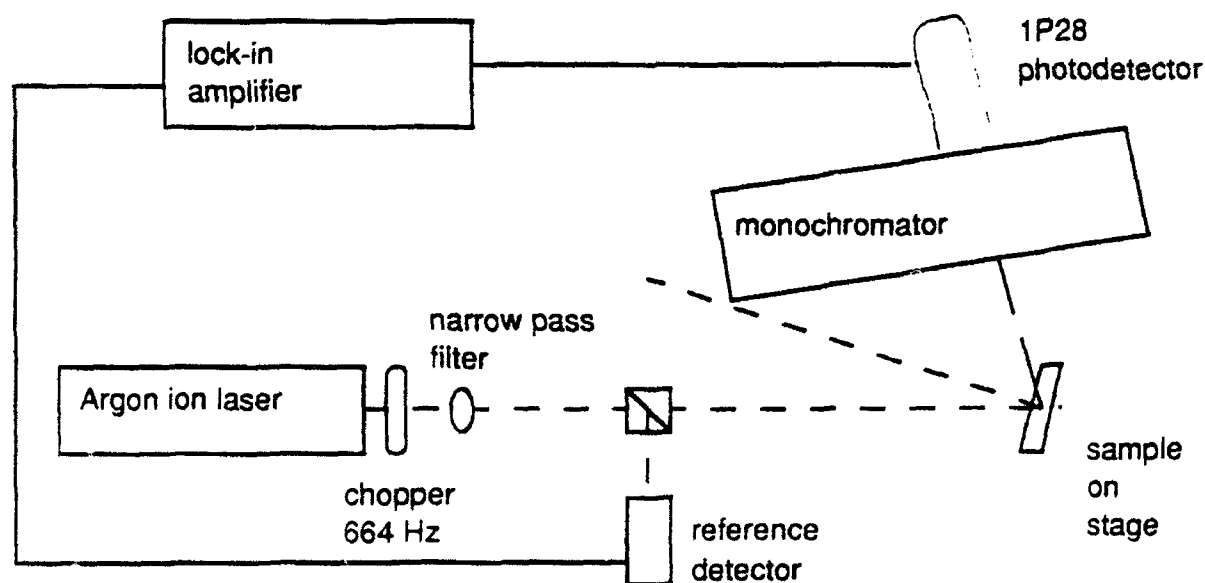


Figure 3. Typical photoluminescence experimental arrangement.

Several authors have examined photoluminescence from GeO_2 [5] and/or $\text{GeO}_2\text{-SiO}_2$ [6-8] glasses. They have identified several features associated with defects. A peak around 400 nm and a broad band in the vicinity of 800 nm are observed in $\text{GeO}_2\text{-SiO}_2$ fibers and in sintered GeO_2 and are both believed to be associated with a Ge^{II} defect. A feature at 530 nm in 500 μm thick sol-gel glass films was attributed to Ge ions but no 400 nm feature was observed. Another feature around 650 nm in fibers was also believed to be related to defects associated with Ge.

We examined films of approximately 6% GeO_2 in SiO_2 which were about 2 microns thick. We were not able to detect any luminescence from these thin films by multiphoton processes. Even when we switched to a pulsed Nd-YAG laser operating in the fundamental or 2nd harmonic modes, we were unable to detect any luminescence. The lack of observed luminescence excited by multi-photon processes may be caused by the very small amount of material available to excite. Previous studies have waveguided through about 1.5 meters of fiber or used bulk or thick film samples. Studies of thin films of other materials sometimes use computerized data acquisition to average over hundreds of experimental runs to pull weak signals out of the background.

In order to increase the amount of film exposed to the incident beam, we tried waveguiding the Ar ion laser beam into the film using a 90° prism to couple them. The microscope objective lens was then aimed at the waveguide region to focus emitted light onto the monochromator entrance slit. Losses associated with coupling the light into the film may have negated the improvements in signal due to increased sample area. Two geometries were tried but no luminescence was detected from either arrangement. One had the lens aimed at the face of the sample over the entire wave guide region. The other turned the sample edge-on to the monochromator and used the lens to collect the light coming out the edge of the sample.

Most photoluminescence experiments use light sources with an energy which is greater than the energy of the transition being observed. This allows direct excitation of the material. It might be better to try a light source with a wavelength in the 300-360 nm range such as the UV lines available from the Ar, Kr, or He-Cd lasers. Although all of these lasers are available at FJSRL/NP the UV optics which is needed in the lasers is not available at this time. Skuja [5] examined the intensity of the photoluminescence as a function of exciting energy and found a maximum for sources with a frequency around 335 nm. The intensity at the Ar laser lines which we used was off the graph but the nearest measured points were near zero. We tried using an arc

lamp and monochromator using both Tungsten-Halogen and Deuterium lamps. We were unable, however, to detect any luminescence most likely because of the relatively low intensity of our light sources and the small amount of material we worked with. We were, incidentally, able to see many other things luminesce using this apparatus - just not our samples.

Only one experiment yielded a possible indication of luminescence. This involved using the Nd-YAG with a quadrupled output (266 nm). Some indication of luminescence was detected around 400 nm from this sample. It was not clear, however, whether the source of luminescence was the film or the substrate. No blank sample of the same type as the substrate was available to run a blank experiment. This experiment was one of the last runs of the summer. It should be noted that several times previously we thought we saw luminescence, but were always able to eliminate it by correcting for artifacts in the experiment. It is possible that this observation was also artifact related.

One potential problem of using UV sources is the possibility of creating defects. This would be an even greater problem if a shorter wavelength (ie. excimer) laser was used. If these UV lines are attempted, the possibility of defect creation should be considered by examining the time dependence of the defect-related luminescence. Another problem is that the optics itself can luminesce giving rise to a spurious signal. (Remember that the sample and the optics are not all that different in composition.)

To improve the experiment, I would recommend using a better monochromator with less stray light (perhaps a double monochromator). This would reduce the background signal which is rather large across the spectrum. A more sensitive detector could then be used, if needed, to try to pick up the small luminescence signal. The use of a better monochromator might eliminate the need for blocking filters which are a possible source of spurious luminescence. The experiment does seem to require extremely high powers of incident light into the sample, even for direct luminescence. The only equipment which would seem to be feasible for this is the quadrupled Nd-YAG laser.

PERIODIC POLING OF THIN FILMS - electric field induced (or assisted) second harmonic generation

FJSRL/NP personnel have been discussing experiments involving externally applying a periodic electric field to induce or assist the second harmonic generation process. I present here two electrode designs and the design parameters for such an experiment. I also model the penetration of the electric field produced into the film.

Kashyap [9,10] has examined the effects of periodically poling fibers on second harmonic generation. Using an interdigitated electrode similar to the one described below, he rotated the electrode to obtain a quasi-phase-matching condition between the electric field periodicity and the incident and second harmonic light. Norwood and Khanarian [11] used fixed electrodes with a polymer waveguide film and varied the incident wavelength to obtain a quasi-phase-matched condition for second harmonic generation. Both authors claim field strengths of $10^6 - 10^7$ V/m from applied voltages of 70 - 120 V.

Electrode design:

Two geometries of electrodes are presented here. One requires an ability to rotate the electrodes with no translation. The other requires translation of the electrodes with no rotation. The choice of electrode design may depend on which of these motions is easier to accomplish and which electrode pattern is easier to have produced.

The fiber studies have used interdigitated electrodes as shown in Figure 4. These depend on rotation of the electrode structure to change the distance between electrodes along the path of interest. As indicated in Figure 4, the pole spacing, λ , can be related to the electrode separation, d , and the angle of rotation from parallel, θ , by

$$\lambda = d/\cos\theta$$

Since d is constant, varying θ will change the pole spacing.

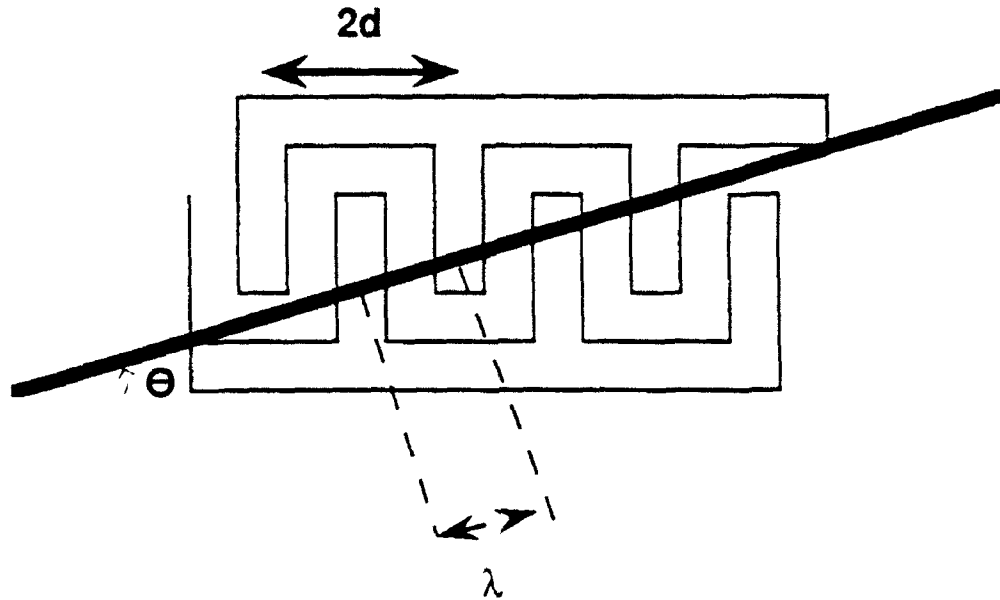


Figure 4. interdigitated electrode design

The other electrode design maintains a constant angular separation between adjacent electrodes and relies on translation of the electrode array to vary the pole spacing. This pattern is shown in Figure 5. We see that the pole spacing relation is now

$$\lambda = L \tan(\phi)$$

where L is the distance from the extrapolated intersection point of the electrodes and ϕ is the constant angular separation between adjacent electrodes.

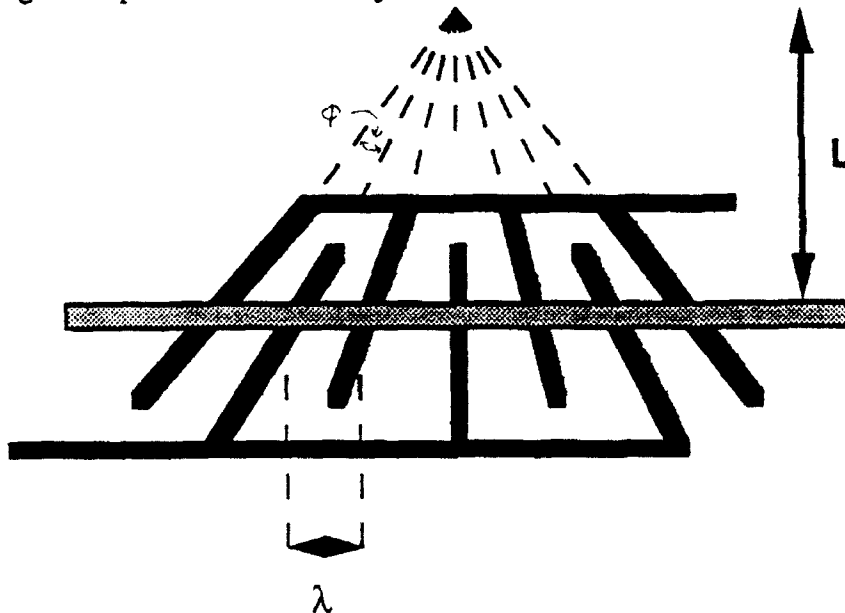


Figure 5. Constant angle electrodes

Typical dimensions of either electrode array would be about 1 cm by 1 cm so that it will fit between the coupling prisms of the waveguide. The electrodes in previous studies were 4 - 16 μm wide and were separated by 20 - 40 μm . The electrodes were metal films (Cr, Al, Au) deposited on a glass (or Si) substrate. A glass substrate is probably easier to work with since you can see through it to confirm the orientation of the electrodes over the wave guide path. If we expect periodic phase matching around 30 μm , we should probably use electrodes separated by around 20 μm in the interdigitated design and 20 - 50 μm in the constant angle design. The electrodes should be about 5 μm wide and about 10 μm thick.

Since the wave guide conditions would be destroyed by having the electrodes in contact with the sample, thin spacers will also have to be deposited on the electrode assembly to hold the electrodes about 1 - 2 μm off of the film surface. It is critical to know the orientation of the wave guide region relative to the electrodes in order to establish the zero for measuring angle or distance.

Using the interdigitated design with a spacing between electrodes of 20 μm , a 21 μm period is achieved by rotating 17.75°. A 30 μm spacing requires 48° of rotation. The electrode array needs to be large enough to allow the waveguide region to pass through a large number of electrodes even at these large rotations. It might be necessary to build several electrodes with different spacings (20, 25, 30 μm) to minimize the amount of rotation needed.

The constant angle design would require electrodes designed with a separation angle of only 0.23° to produce a 20 μm separation between electrodes at a distance of 5 mm from the intersection point. This angle would be 1.15° if the 20 μm spacing was achieved only 1 mm from the intersection. For the 0.23° spacing, a 1 micron change in electrode spacing would be achieved by a 250 μm translation of the electrodes. For the 1.15° angle, this translation would be 50 μm .

Field magnitude and orientation:

The nature of the electric field inside the film can be approximated by examining the case of two adjacent semi-infinite plates separated by a distance, $2a$, and having potentials V_1 and V_2 respectively ($V_2 > V_1$). This configuration is shown in Figure 6.

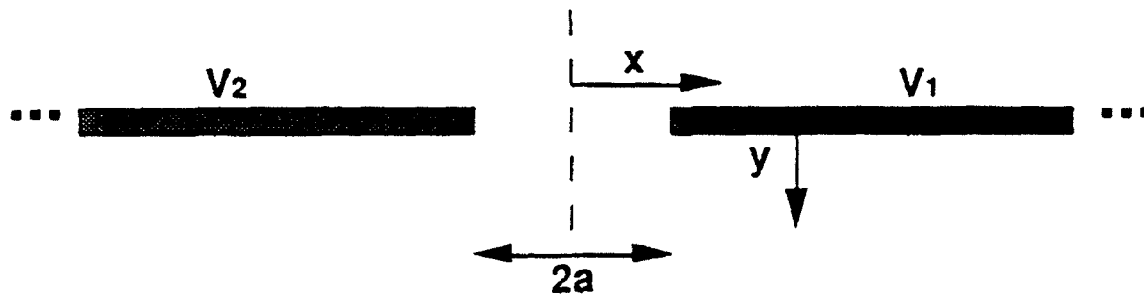


Figure 6. Adjacent semi-infinite plates

Using a Schwarz-Christoffel transformation, this problem can be solved [12,13] to yield a complex potential function

$$F - iV = (1/\pi) (V_2 - V_1) \cosh^{-1} (z/a) - iV_1$$

where F is the flux function, V is the potential and $z = x + iy$ with (x,y) being the spatial coordinates of the potential and flux. Using the identity

$$\cosh^{-1} z = \log (z + (z^2 - 1)^{1/2})$$

we can find the potential from the imaginary part of the expression and the flux from the real part.

I have written a simple FORTRAN program on the VAX which performs these calculations for any spacing of the plates and potential. The program is included in Appendix A. Using this program, we can estimate the effects of plate separation and applied potential on the direction and magnitude of the electric field.

We first vary the plate separation from 20 to 40 μm with a fixed potential of ± 100 V on the plates. The direction of the field is only slightly more vertical for the larger separation. Remember that in the real case of a repeating electrode pattern, the field at the center of the electrode will be constrained by symmetry to be vertical - regardless of spacing. The magnitude of the electric field can be estimated from the gradient of the potential. Assuming a dielectric constant of 3.8 for fused SiO_2 and neglecting the air gap between the electrodes and film, the field at a point 2 μm below the plates and 2 μm in from the edge decreases from 1.9×10^6 V/m to 1.4×10^6 V/m when the separation increases from 20 to 40 μm .

Changing the plate potentials from ± 100 V to ± 50 V for a fixed separation of 20 μm again changes the field direction minimally. The electric field magnitude at a point 2 μm below the

plates and 2 μm in from the edge of the plates is $1.9 \times 10^6 \text{ V/m}$ for $V = \pm 100 \text{ V}$ and $0.95 \times 10^6 \text{ V/m}$ for $V = \pm 50 \text{ V}$. The only limit on increasing the potential on the electrodes would seem to be the breakdown strength of the dielectric film. For SiO_2 , this value is $6 - 7 \times 10^8 \text{ V/m}$ at room temperature, so this is not a significant constraint.

The electric field magnitude is found to decrease as you go into the film (away from the electrode plate) as expected. Figure 7 indicates the nature of this decrease for a field in fused SiO_2 with electrode potentials of $\pm 100 \text{ V}$, a spacing between plates of $20 \mu\text{m}$, and at a distance 2 μm in from the edge of a plate.

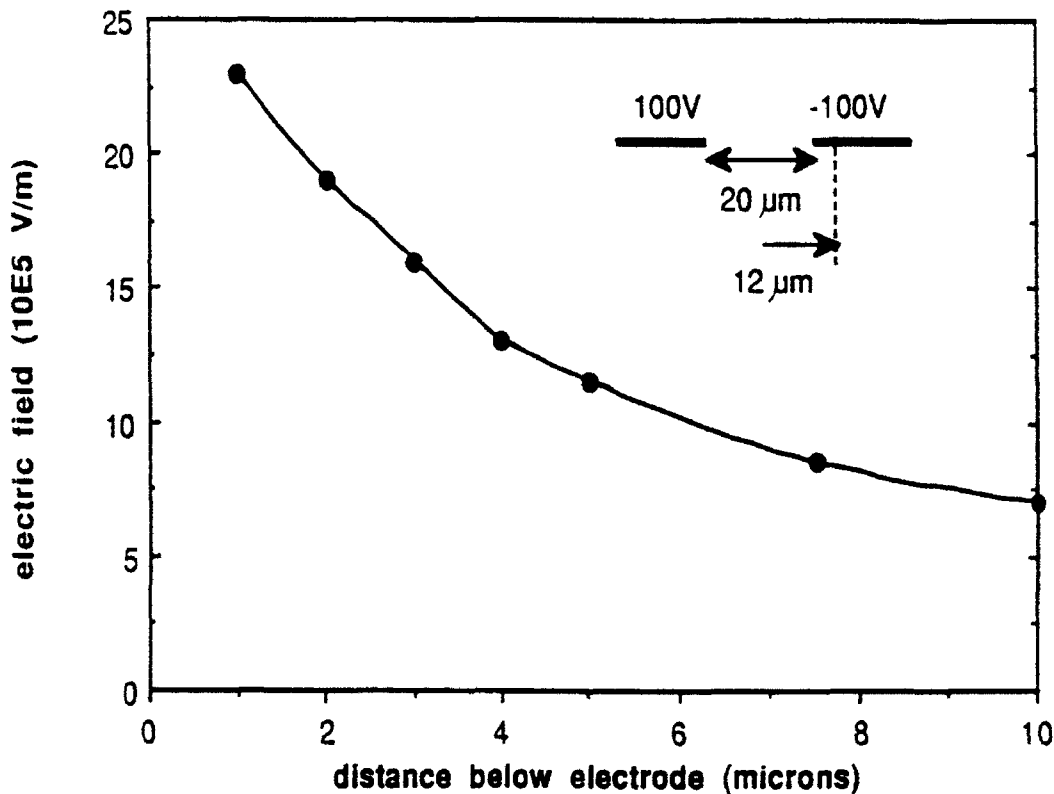


Figure 7. Variation of electric field strength with distance below the electrode at a point 2 μm from the edge of the plate.

SECOND HARMONIC GENERATION FROM ACTIVATED REGIONS

In waveguiding in thin films only small paths across the film are made non-linearly active. These regions are approximately 25 μm wide and extend across the sample. Using a Maker fringe apparatus, I attempted to detect a second harmonic generation from these thin film samples.

The samples were placed in the beam path of a Nd-YAG laser which was focused down to about 10 μm at the sample. The average power at the sample was typically 100-200 mWatts. A detector was placed behind the sample and filtered to detect only second harmonic signals from the sample. The sample was set at a fixed angle of 30, 60, or 90° with respect to the incident beam. We hoped to detect second harmonic generation if the beam struck a non-linearly active region of the sample and then hoped to observe the angular dependence of the second harmonic signal. This required the film to be directly on the axis of rotation so that rotating the sample would not move it out of the beam.

The dimensions involved were all on micron scales resulting in a need for very careful positioning and presumably resulting in very small second harmonic signals. None of our experiments yielded any detectable second harmonic light. We varied the polarization of the incident beam as well with no success.

The two most likely explanations for our lack of results are insufficient sensitivity to very small second harmonic signals or lack of fine enough positioning control to hit a non-linearly active region of the sample. The first problem may require very sensitive photon counting detectors to correct. Computer controlled translation stages with about one micron steps might help reduce uncertainties from manually operated translation. Combined with computer control of the detector to allow better averaging, the experiment could be completely automated.

REFERENCES:

1. U. Osterberg and W. Margulis "Dye laser pumped by Nd:YAG laser pulses frequency doubled in a glass optical fiber" *Opt. Lett.* 11, 516 (1986).
2. R. H. Stolen and H. W. K. Tom "Self-organized phase-matched harmonic generation in optical fibers" *Opt. Lett.* 12, 585 (1987).
3. J. Kester et al. *Opt. Lett.* submitted
4. L.E. Davis et al., "Handbook of Auger Electron Spectroscopy" 2nd ed., (Physical Electronics, Eden Prairie, MN, 1976).
5. L. N. Skuja "Photoluminescence of intrinsic defects in glassy GeO_2 " *Phys. Stat. Sol. (a)* 114, 731 (1989).
6. Y. Hirai, T. Fukuda and K. Kubota "Optical properties of SiO_2 - GeO_2 glasses made by sol-gel method" *J. Non-Crystalline Solids* 93, 431 (1987).
7. B. Poumellec, J. M. Gabriagues and H. Fevrier "Anomalous fluorescence-like generation in silica-based optical fibres" *J. Phys. Chem. Solids* 51, 1397 (1990).
8. B. Poumellec, H. Fevrier and J.-M. Gabriagues "Defect -induced non-linear optical properties of silica-based optical fibres" *Materials Sci. and Eng.* B9, 449 (1991).
9. R. Kashyap "Phase-matched periodic electric-field-induced second harmonic generation in optical fibers" *J. Opt. Soc. Am. B* 6, 313 (1989).
10. R. Kashyap "Phase-matched second-harmonic generation in periodically poled optical fibers" *Appl. Phys. Lett.* 58, 1233 (1991).
11. R. A. Norwood and G. Khanarian "Quasi-phase-matched frequency doubling over 5 mm in periodically poled polymer waveguide" *Electronics Lett.* 26, 2105 (1990).
12. E. Weber, "Electromagnetic Fields: Theory and Applications: Volume 1 - Mapping of Fields", (John Wiley and Sons, Inc., New York, 1950) p. 339.
13. K. J. Binns and P. J. Lawrenson, "Analysis and Computation of Electric and Magnetic Field Problems", 2nd ed. (Pergamon Press, Oxford, 1973) p. 171.

Appendix A: FORTRAN program

```

C ELECEN.FOR
C
C THIS PROGRAM WILL CALCULATE THE POTENTIAL AND FLUX FROM TWO
C ADJACENT SEMI-INFINITE PLATES.
C
C Tom Christensen, EUEL/NP
C July 9, 1992
C VAX - FORTRAN
C
C COMPLEX*8 Z, Z1, Z2
C
C TYPE*, ' '
5 TYPE*, 'THIS PROGRAM CALCULATES POTENTIAL AND FLUX FROM'
TYPE*, 'TWO ADJACENT SEMI-INFINITE PLATES'
TYPE*, 'The program has two modes:'
TYPE*, '1. calculate flux and potential at a given point'
TYPE*, '2. calculate lines of constant flux'
TYPE*, 'ENTER MODE DESIRED:'
ACCEPT*, INODE
C
7 TYPE*, ' '
TYPE*, ' _V2_____ : _____V1_'
TYPE*, '          ~~~~~ A'
TYPE*, ' V2/V1          :----->X'
TYPE*, ' '
C
TYPE*, 'ENTER SPACING BETWEEN PLATES (2A):'
ACCEPT*, SPACE
A = SPACE/2
C
TYPE*, 'ENTER POTENTIAL ON PLATE 1 (V1(V2)):'
ACCEPT*, V1
TYPE*, 'ENTER POTENTIAL ON PLATE 2 (V2(V1)):'
ACCEPT*, V2
C
IF (INODE .EQ. 2) GO TO 46
C
18 TYPE*, 'ENTER X DISTANCE FROM MIDPOINT BETWEEN PLATES:'
ACCEPT*, X
TYPE*, 'ENTER Y DISTANCE FROM PLATES:'
ACCEPT*, Y
Z = CMPLX(X,Y)
C
Z1 = Z/A
Z2 = CLOG(Z1 + CSQRT(Z1**2 -1))
C
VR1 = AIMAG(Z2)
FR1 = REAL(Z2)
C
POTL = ((V2-V1)/3.14159)*ABS(VR1) + V1
FLUX1 = ((V2-V1)/3.14159)*ABS(FR1)
TYPE*, 'potential=', POTL, 'flux=', FLUX1
C
TYPE*, 'CHANGE PLATE PARAMETERS (Voltage- separation)?'
ACCEPT*, IDO
100 FORMAT(1A2)
IF((IDO .EQ. 'Y') .OR. (IDO .EQ. 'y')) GO TO 7

```

```

C
type= 'Change spatial position?'
accept 100= IDO
if((IDO .eq. 'Y') .or. (IDO .eq. 'y')) goto 15
C
      goto 99
C
C Mode 2:
C Calculate initial flux value at plate
C
49 type= 'Enter x distance from midpoint between plates:'
accept= X
Y = 0.0
Z = cmplx(X,0.0)
ZI = Z/A
ZZ = CLOG(ZI + CSQRT(ZI**2 -1))
C
VR1 = AIMAG(ZZ)
FR1 = REAL(ZZ)
C
POTL0 = ((V2-V1)/3.14159)*ABS(VR1) + V1
FLUX00 = ((V2-V1)/3.14159)*ABS(FR1)
type= X, ' 0.000000P, FLUX00'
C
C Iteratively solve for (x,y) values having the same flux.
C Note solution spacing is fixed in the program as steps of 1 unit
C in the x direction and 0.003 units in the y direction.
C
49 X = X - 1
Y = Y - 0.003
59 Y = Y + 0.003
Z = cmplx(X,Y)
ZI = Z/A
ZZ = CLOG(ZI + CSQRT(ZI**2 -1))
C
VR2 = AIMAG(ZZ)
FR2 = REAL(ZZ)
C
POTL = ((V2-V1)/3.14159)*ABS(VR2) + V1
FLUX0 = ((V2-V1)/3.14159)*ABS(FR2)
C
ERROR = ABS(FLUX00 - FLUX0)
if (ERROR .gt. 0.1) goto 59
type= X, Y, FLUX0
if (X .le. 0) goto 69
goto 49
69 type= 'Calculate flux line at another point?'
accept 100= IYES
if ((IYES .eq. 'Y') .or. (IYES .eq. 'y')) goto 49
C
type= 'CHANGE PLATE PARAMETERS (Voltage separation)?'
ACCEPT 100= IDO
IF((IDO .eq. 'Y') .or. (IDO .eq. 'y')) GO TO 7
C
99 stop
end

```

Modeling of the NLO properties of Silica

Ravi Pandey
Physics Department, Michigan Tech University
Houghton, MI 49931.

Final Report for :
Summer Research Program
Frank. J. Seiler Laboratory

Sponsored by :
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D. C.

August 1992

Modeling of the NLO properties of Silica

Ravi Pandey
Physics Department, Michigan Tech University
Houghton, MI 49931.

ABSTRACT

The microscopic NLO properties in terms of static and frequency-dependent polarizabilities of pure and Ge-doped silica glasses are determined using a cluster approximation. Both the semi-empirical and ab initio methods are used to describe $\text{Si}(\text{OH})_4$ and $\text{Ge}(\text{OH})_4$ clusters. The semi-empirical PM3 model appears to provide a reliable starting point for ab initio calculations. The calculated results show that the cluster configuration induced by the presence of a defect plays a significant role in predicting a non-linear response of the silica glass.

Modeling of the NLO properties of Silica

Ravi Pandey

I. INTRODUCTION :

Nonlinear optics is currently at the forefront of research because it is the key technology for optical communication, signal processing and computing.¹ Photonic switches made out of nonlinear optical (NLO) materials have potential advantages of wideband frequency domain, immunity to electromagnetic induction, and high speed operation over electrical switches. Glassy materials are one of the prime candidates for Photonic switches due to there (1) ease of fabrication, (2) high transparency, (3) fast switching time, and (4) high chemical and thermal stability.

Nonlinear optical properties of glasses are known to be correlated with the third order susceptibility due to the absence of ordered dipoles. However, Kester et al.² have observed the frequency doubling in Ge-doped silica glasses suggesting that even glasses have frequency doubling capability. Similar observation has been made in bulk and thin films of SiO₂ grown on semiconductor substrates.³

As part of ongoing research in AFOSR/FJSRL-NC in the Nonlinear Optics area, the present work attempts to provide a basis to understand the origin of such NLO effects in pure and impurity-doped silica glasses at the molecular level. We report here the results of a preliminary study of the silica glass that is simulated by a molecular cluster treated in either semi-empirical or *ab initio* methods. Our efforts will be directed to polarizabilities' calculations that are related to the microscopic NLO properties.

In the following section, we briefly describe the computational model. The results are presented and discussed in Sec. III and the recommendations for the future work are given in Sec. IV.

II COMPUTATIONAL MODEL :

The various forms of silicon dioxide (either crystalline or amorphous) are known to share a similar short-range order⁴, consistent with the chemistry of the Si-O bond : each Si atom is coordinated (nearly) tetrahedrally with four O neighbors, and each O atom has two Si neighbors with a Si-O-Si angle of about 150°. The relatively flexibility of Si-O-Si / O-Si-O angle and the dihedral angle associated with adjacent SiO₄ tetrahedra allows the formation of the variety of structures of silica.

Our computational model simulates the silica glass (pure or imperfect containing defects such as Ge) by a cluster of atoms with appropriate boundary conditions assuming that cluster wave function is well localized. In order to suppress the so-called surface effects arising from the dangling bonds of outermost atoms⁵, we use hydrogen atoms. Each oxygen atom at the cluster boundary is attached by a hydrogen atom in such a way that they simulate O-Si bond directions.

In this preliminary work, we limit ourselves to 9-atom clusters, namely Si(OH)₄ and Ge(OH)₄. The optimum configurations of these 9-atom clusters are obtained using both semi-empirical (MOPAC-AM1/PM3) and ab initio (GAMESS-ECP/DZV) approaches expecting that MOPAC configuration will provide a reliable starting point for more computationally expensive GAMESS calculations.

MOPAC is a general-purpose, semi-empirical molecular program implementing the semi-empirical Hamiltonians MNDO, AM1, PM3 and MNDO/3.⁶ On the other hand, GAMESS is a general atomic and molecular electronic structure program.⁷ For this work we use the Hartree-Fock approximation where the cluster atoms are described with double zeta valence (DZV) Gaussian basis sets. We also use the effective core potentials given by Stevens, Basch and Krauss (ECP) for the cluster atoms to see their effects on the all-electron (DZV) results.

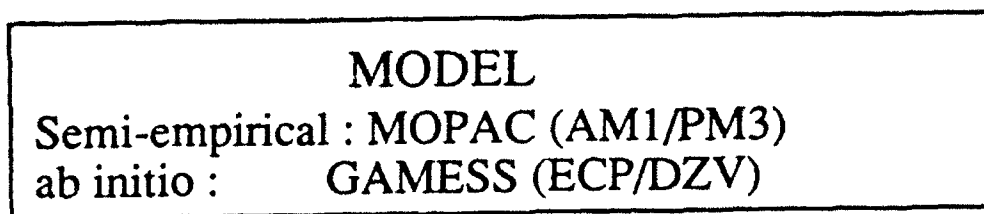
We note here that the nonlinearity effect in a dielectric medium may be of microscopic or macroscopic origin. The polarization density $P = N \mu$ is a product of individual dipole moment μ and, which is induced by the applied electric field F , and the number of density of dipole moments N . The nonlinear behavior may have its origin in either μ or in N .

For the dipole moment of a molecule interacting with a static electric field (F), we can write :

$$\mu_i = \mu_{i0} + \alpha_{ij} F_j + (1/2) \beta_{ijk} F_j F_k + (1/6) \gamma_{ijkl} F_j F_k F_l + \dots \quad (1)$$

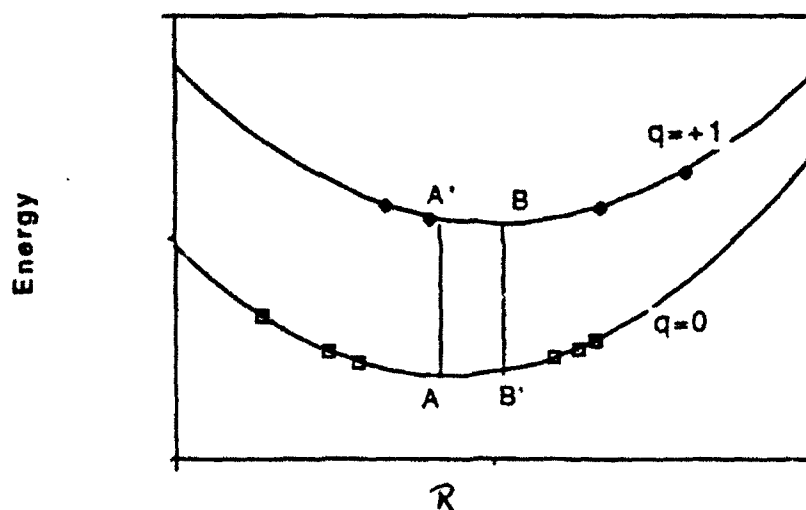
where μ_{i0} is the permanent dipole moment and α_{ij} , β_{ijk} and γ_{ijkl} are tensor elements of the linear polarizability, and first and second hyperpolarizabilities, respectively of the molecule.⁸ Likewise, one can write an expression for dynamic fields relating β and γ to experiments such as second harmonic generation, optical rectification, and Kerr effect.⁹

As shown in Fig. 1, the potential energy surfaces and then subsequently polarizabilities at the optimum configuration of both neutral ($q=0$) and singly charged (ionized) states ($q=+1$) of the X(OH)₄ cluster (X being Si or Ge) are calculated. The $q=+1$ charge state is expected to mimic the defect-induced effect on the NLO properties. We also



Geometry Optimization

NLO properties
(a, b, g) Finite Field Calculations
Static / Frequency Dependent (ACES II)



use the vertical approximation for calculations of the single point A' (geom A, $q=+1$) and B' (geom B, $q=0$).

III. RESULTS and DISCUSSION :

Table 1 shows the geometric parameters, namely the bond-length and the bond-angles of the cluster optimum configuration. The semi-empirical Hamiltonian, AM1 appears to overestimate the bond-length in either $\text{Si}(\text{OH})_4$ or $\text{Ge}(\text{OH})_4$ cluster. The optimum configuration obtained using either ECP's or DZV basis set turns out to be the same. When we substitute Ge for Si, the bond-length increases in all the cases, as expected. Further the Ge is shown to be less electronegative than Si in these clusters. For the ionized cluster ($q=+1$), the optimum configuration shows the hole localizing on one of the neighboring Oxygens resulting in an increase of its bond-length (Table 2). We note here that PM3 calculation predicts the hole localization on two of the neighbor Oxygens in contrast to AM1, ECP and DZV calculations for $\text{Si}(\text{OH})_4$ cluster.

In order to judge the reliability of predicted optimum configuration, we calculate the frequency of the $\text{Si}(\text{OH})_4$ cluster in the harmonic approximation and compare with the experimentally known frequencies. As shown in Table 3, the calculated frequencies in rocking, bending and stretching modes in both ECP and DZV model show a good agreement with the corresponding experimental values. Furthermore, the total electronic density of $\text{Ge}(\text{OH})_4$ cluster is shown in Fig. 2 indicating a predominant covalent bond formation in the cluster, as expected.

Table 4 lists the calculated polarizabilities of neutral $\text{Si}(\text{OH})_4$ and $\text{Ge}(\text{OH})_4$ clusters at the optimum configuration in the semi-empirical (AM1 and PM3) and ab initio (ECP and DZV) methods. It appears that all the models agree in predicting the trend for α and β when we substitute Ge for Si. However, the semi-empirical methods disagree with ab initio methods for γ . In the following, we therefore give results obtained using GAMESS-DZV.

The NLO properties have been shown to very sensitive to the choice of basis-set. For the case of ethylene, a sufficiently flexible valence Gaussian basis set, such as a 3-21G or 6-31G, augmented by diffuse p and d functions on C atoms is found to be necessary for a satisfactory description of various tensor components of static polarizability and second hyperpolarizability.¹⁰ Similarly, an extended, even-tempered diffuse basis-set is found to provide an adequate description of all the components of the first hyperpolarizability for H_2O .¹¹ We therefore add a diffuse d-type function for the polarizabilities calculation to the optimum configuration obtained at DZV level.

X(OH)₄ cluster :

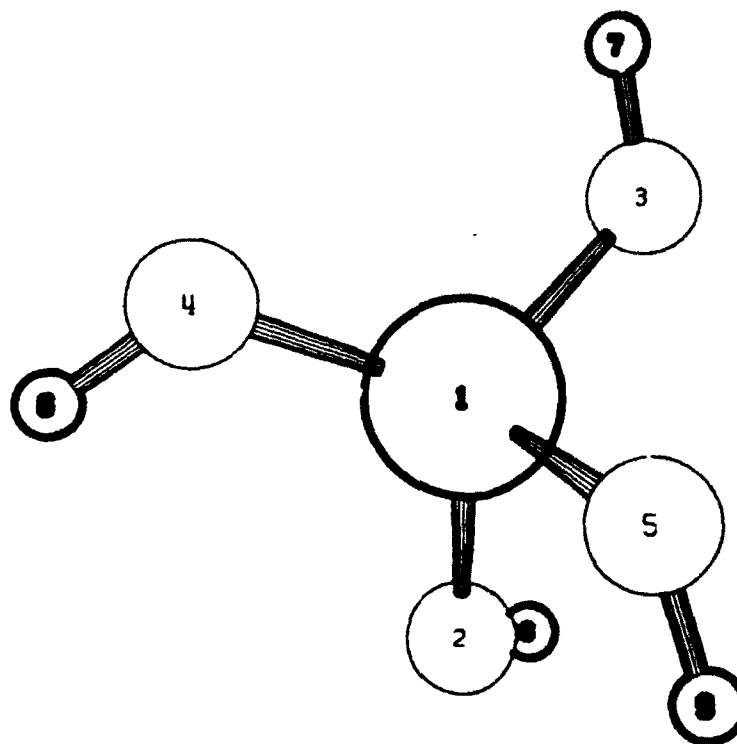


Table 1

		MOPAC		GAMESS	
		AM1	PM3	ECP	DZV
Bond-length(Å)					
Si-O		1.728	1.685	1.655	1.652
Ge-O		1.868	1.798	1.721	1.726
Bond-angle(°)					
O(2)-Si-O(i), i=3,5		105.1, 118.7	107.8, 112.8	107.2, 114.1	107.3, 113.8
O(2)-Ge-O(i), i=3,5		107.1, 114.4	107.9, 112.7	106.8, 115.0	107.6, 113.3
Charge					
-Si		+1.69	+1.20	+1.59	+2.26
-Ge		+1.51	+0.94	+1.51	+2.18

Table 2

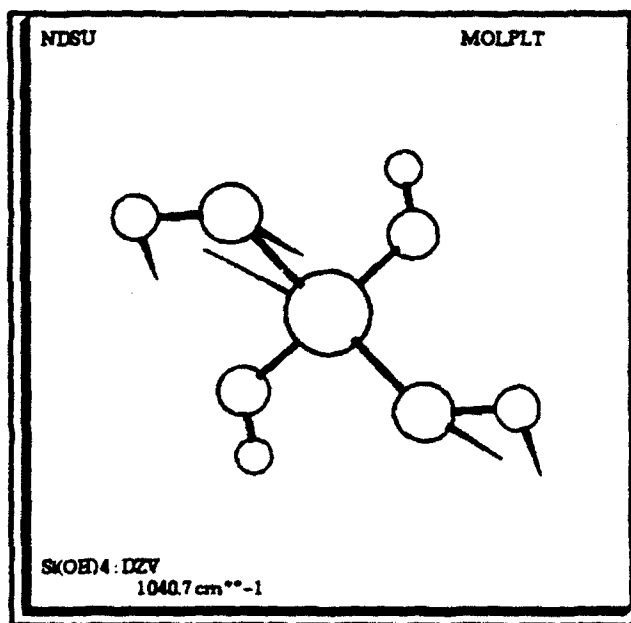
Configuration of neutral and ionized Si(OH)_4 and Ge(OH)_4 clusters.

	R(X-O(i), i=2,5) (Å)	A(O(2)-X-O(i), i=3,5)
AM1 :		
Si : q=0	1.728, 1.728, 1.728, 1.728	105.1, 118.7, 105.1
q=+1	1.694, 1.685, 1.685, 1.978	117.6, 122.3, 79.2
Ge : q=0	1.868, 1.868, 1.868, 1.868	107.1, 114.4, 107.1
q=+1	1.839, 1.841, 1.843, 2.021	116.9, 119.9, 77.4
PM3 :		
Si : q=0	1.685, 1.685, 1.685, 1.685	107.8, 112.9, 107.8
q=+1	1.640, 1.706, 1.640, 1.706	112.0, 126.2, 112.2
Ge : q=0	1.798, 1.798, 1.798, 1.798	107.9, 112.7, 107.9
q=+1	1.767, 1.767, 1.767, 1.901	117.3, 118.9, 95.9
GAMESS-ECP :		
Si : q=0	1.655, 1.655, 1.655, 1.655	107.2, 114.1, 107.2
q=+1	1.588, 1.597, 1.591, 1.980	117.8, 119.7, 96.2
Ge : q=0	1.721, 1.721, 1.721, 1.721	106.8, 115.0, 106.8
q=+1	1.666, 1.663, 1.664, 2.081	118.6, 119.8, 92.8
GAMESS-DZV :		
Si : q=0	1.652, 1.652, 1.652, 1.652	107.3, 113.8, 107.3
q=+1	1.594, 1.600, 1.584, 1.943	117.8, 119.4, 95.6
Ge : q=0	1.726, 1.726, 1.726, 1.726	107.6, 113.3, 107.6
q=+1	1.655, 1.653, 1.654, 1.953	117.7, 118.6, 96.4

Table 3

Si(OH)₄ :Normal Coordinate Analysis in the Harmonic Approximation.
(Frequencies in cm⁻¹)

MODE	ECP	DZV	EXPT.
rocking	441	465	450
bending	808	828	812
stretching	1018	1041	1058



Ge(OH)4-DZV: <Plane-3ATOMS:4 3 2>

DENDIF

ONE BOHR IS —

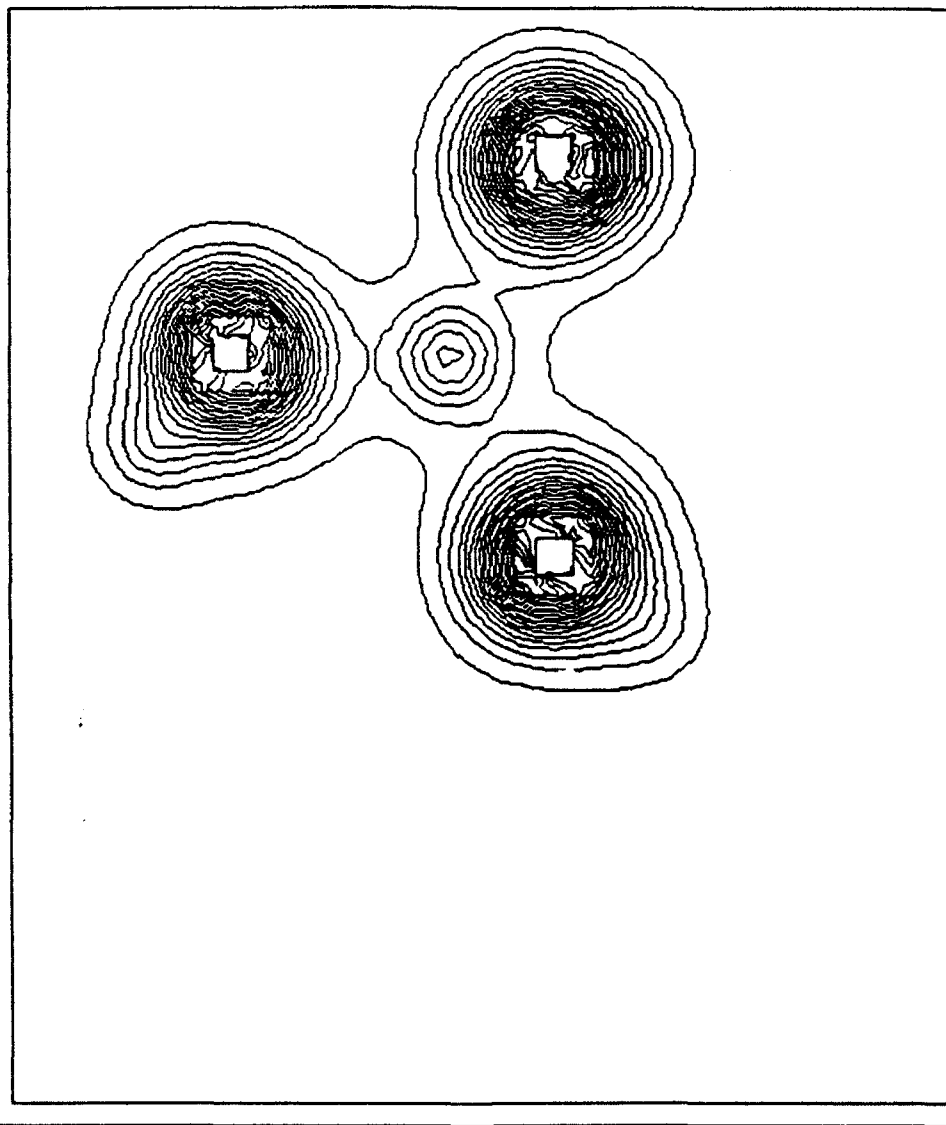


Table 4

Finite Field Calculations of neutral X(OH)₄ cluster.

		MOPAC		GAMESS	
		AM1	PM3	ECP	DZV
α	($\times 10^{-24}$ esu) : Si	7.87	6.53	7.47	7.29
	Ge	8.27	8.13	8.26	8.25
β	($\times 10^{-30}$ esu) : Si	0.001	0.00003	0.00064	0.00028
	Ge	0.001	0.00001	0.00046	0.00023
γ	($\times 10^{-36}$ esu) : Si	0.023	1.90	0.819	0.525
	Ge	0.205	0.983	0.993	0.679

Table 5 lists the value of Dipole moment, β and γ for Si(OH)_4 and Ge(OH)_4 clusters. As we ionize the cluster (freezing the configuration at $q=0$), the β shows a large, noticeable increase with an increase of cluster dipole moment and γ also. The optimized configuration (for $q=+1$) retains the somewhat large β value. However, when we make the cluster charge to be neutral, the dipole moment, β and γ show again an increase of their values. Similar observation has been made for Ge(OH)_4 cluster except the case for $q=+1$ (at $q=0$ configuration). For this case, there is a large increase of γ with no change in dipole moment.

It appears therefore that the cluster configuration plays a key role in predicting a large β value that is related to second-harmonic generation or Kerr effect. The results show a dependence of β on the cluster dipole moment resulting from the distorted configuration. However the calculated value of γ does not show such dependence.

It has been shown that the dispersion of NLO effects is of great importance for the application of conjugated polymers to photoionic studies. But this is found to be not the case for both Si(OH)_4 and Ge(OH)_4 clusters. We give the results of time-dependence Hartree-Fock (TDHF) calculations using the program package ACES II¹² at 1064 nm in Table 6. Although the frequency-dependent values are predicted to be about the same as static values, they do confirm the role of configuration for a large β value.

IV RECOMMENDATIONS :

In the case of organic materials, the origin of non-linear optical responses is attributed to either the existence of low-lying electronic state with a large dipole moment, or through highly polarizable lone pairs of electrons.¹³ For the present case, the cluster configuration induced by the presence of a defect seems to determine the microscopic chemical mechanism of the NLO effect.

Since the identification and structural models of most of the defects in SiO_2 have been achieved by EPR, the geometric and electronic structures of this diamagnetic defect have remained largely speculative. For the future work, we propose to enlarge the cluster size to $\text{Si(OSi(OH)}_3)_4$ so that an accurate role of defects such as oxygen vacancies and different charge states of the impurity Ge on the NLO properties of silica glasses can be determined.

Table 5

Finite Field Calculations (GAMESS) of neutral and ionized Si(OH)_4 and Ge(OH)_4 clusters (in atomic units).

		Dipole	Beta	Gamma
Si :	q=0	0.00	0.03	1040.0
	q=+1	1.41	52.7	2430.9
	q=+1, opt	0.24	10.6	1442.2
	q=0	1.58	61.8	1689.3
Ge:	q=0	0.00	0.03	1345.1
	q=+1	0.00	11.7	38123.2
	q=+1, opt	1.35	38.7	2323.0
	q=0	0.92	15.5	1608.5

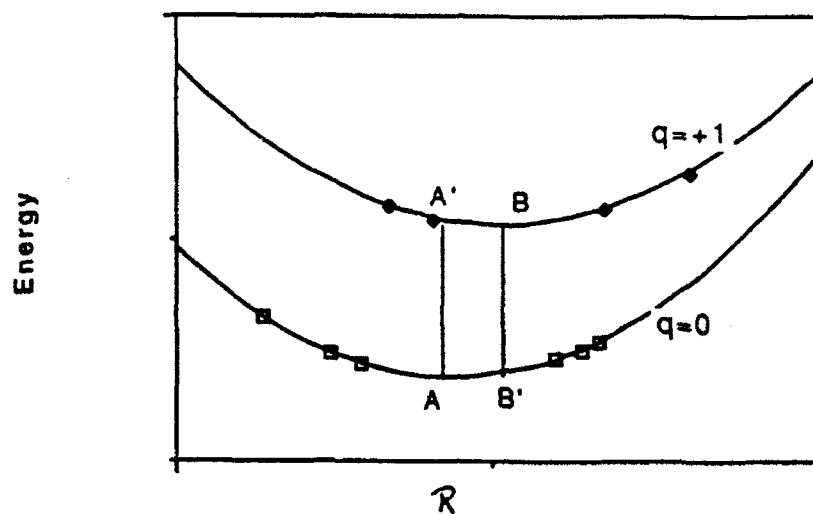


Table 6

TDHF results (at 1064 nm) in atomic units : $\text{Si}(\text{OH})_4$

		geom A	geom B'
BETA :	Static	0.11	60.3
	SHG	0.11	63.8
	Kerr effect	0.11	61.5
GAMMA :	Static	1122.3	1593.2
	DC-SHG	1185.7	1713.4
	THG	1254.5	1848.7

ACKNOWLEDGEMENTS :

I would like to thank the personnel of FJSRL and USAFA for help received during the research performed here. Particular thanks are due to Capt. M. Coolidge who helped me tremendously with the acquisition of the knowledge about MOPAC and GAMESS, and Capt. W. Lauderdale who helped me learn about TDHF and ACES. Also, I must express my appreciation to all the civilian and military personnel at FJSRL, particularly D. Bagely for her help.

I wish to thank the Air Force Materials Command, the Air Force Office of Scientific Research, and Division of Chemical Sciences of the Seiler Research Laboratories at the USAF Academy for sponsorship of this research, and Research Development Laboratory for their assistance in the administrative aspects of the program.

REFERENCES :

- ¹J. D. Mackenzie, C. Y. Li and Y. Xu, Chemistry Express, 6, 903 (1991).
- ²J. Kester et al., private communication.
- ³R. A. Myers, N. Mukherjee and S. R. Brueck, Nonlinear Optics: Topical Meeting, August 17-21, 1992, Hawaii, abstract MA3.
- ⁴F. L. Galeener, in The Physics and Technology of Amorphous SiO₂, ed: R. A. B. Devine (Plenum, New York, 1988), pp. 1-13.
- ⁵For a detailed discussion, we refer to J. K. Rudra and W. B. Fowler, Phys. rev. B 35, 8223 (1987).
- ⁶J. P. Stewart, MOPAC : a general molecular orbital package, Quantum Chemistry Program Exchange, No. 455, (1983), Version 6 (1990).
- ⁷M. W. Schmidt, K. K. Baldridge, J. A. Boatz, J. H. Jensen, S. Koseki, M. S. Gordon, K. A. Nguyen, T. L. Windus, S. T. Elbert, QCPE Bulletin, 10, 52 (1990).
- ⁸C. Flytzanis in Theory of Nonlinear Optical Susceptibilities, Quantum Electronics, H. Rabin and C. L. Tang, Eds., Academic Press, New York, 1975.
- ⁹D. S. Chemla and J. Zyss, Eds., Nonlinear Optical Properties of Organic Molecules and Crystals Vol. I & II, Academic Press, New York, 1987.
- ¹⁰S. P. Karna, G. B. Talapatra, W. M. K. P. Wijekoon and P. N. Prasad, Phys. Rev A 45, 2763 (1992)
- ¹¹W. A. Parkinson and J. Oddershede, J. Chem. Phys. 94, 7251 (1991).
- ¹²J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale and R. J. Bartlett, University of Florida, 1991.
- ¹³Nonlinear optical properties of Organic Molecules and Crystals, edited by D. S. Chemla and J. Zyss (Academic, New York, 1987).

ELECTROCHEMISTRY IN LITHIUM CHLORIDE BUFFERED-NEUTRAL
ROOM TEMPERATURE MELTS

Bernard J. Piersma
Professor
Department of Chemistry

Houghton College
Houghton, N.Y. 14744

Final Report for:
Summer Research Program
Frank J. Seiler Research Laboratory,
USAF Academy, CO 80840-6528

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

August 1992

Electrochemistry in Lithium Chloride Buffered-Neutral
Room Temperature Melts

Bernard J. Piersma
Professor
Department of Chemistry
Houghton College

Abstract

The feasibility of using room temperature chloroaluminate molten salts for the development of high energy-density batteries has been the focus of on-going electrochemistry research at this laboratory. The primary objective of this project is to determine whether lithium would be a suitable anode for a secondary battery in 1-methyl-3-ethylimidazolium chloride chloroaluminate molten salt electrolyte. Initially several approaches were explored to achieve the buffering of neutral melts (neutral in terms of Lewis acidity) using lithium chloride as the buffering agent. Following this the role of protons, added as methylethylimidazolium hydrogendifluoride, to facilitate lithium deposition and stripping and to extend the melt window was studied using tungsten, platinum, gold and glassy carbon electrodes. Finally the stability of lithium in an optimized proton-rich buffered-neutral melt and the charge-discharge processes that lithium would undergo as the anode in a secondary battery were studied using cyclic voltammetry, chronoamperometry and other electrochemical techniques.

Electrochemistry in Lithium Chloride Buffered-Neutral
Room Temperature Melts

Bernard J. Piersma

INTRODUCTION

Room Temperature molten salts formed by mixing 1-methyl-3-ethylimidazolium chloride (MEIC) with aluminum chloride have been intensively studied at FJSRL for more than a decade with the goal of developing high energy density batteries. Alkali metals have high oxidation-reduction potentials and relatively low atomic masses making them very attractive candidates as battery anodes. Lithium is of particular interest because it has the highest electricity storage density of the active metals. Thus the demonstration of a stable reversible lithium anode would be an important step in the development of practical secondary batteries using these chloroaluminate melts as electrolytes.

The Lewis acid-base behavior of these melts has a substantial influence on their electrochemical and physical properties. With MEIC as donor of the Lewis base chloride ion and AlCl_3 as the Lewis acid, the following reactions apply:



The melt is Lewis basic when chloride ion is present in excess (excess (MEIC) and Lewis acidic when AlCl_3 is present in excess (Al_2Cl_7^- is present). A melt is neutral when the only aluminum species present is AlCl_4^- . Melt acidity is easily ascertained electrochemically since Al deposition only occurs from acidic melts and the melt window (region of electrochemical stability) is significantly greater in neutral melts than in either basic or acidic melts. Because of the wide electrochemical window, it is desirable to work with a melt which is exactly neutral, but as with most acid-base systems maintaining neutrality is not easily achieved unless the system is buffered. Fortunately it was recently determined

that this melt system can be easily buffered to the neutral composition using sodium chloride(1).

A second important observation critical to the development of reversible alkali metal anodes in buffered-neutral melt is that protons added to the melt as MEIHC_l, provides a more negative voltage window which significantly reduces the reactivity of the active metal with the melt(2). This study demonstrated that nearly reversible deposition-stripping behavior could be achieved in the buffered-neutral melt with sodium when the melt was made sufficiently rich in proton concentration. Our intent for this project is to exploit these important discoveries and determine whether the lithium redox couple is suitable as the anode for a rechargeable battery using a proton-rich room temperature molten salt buffered to neutral Lewis acidity with lithium chloride.

METHODOLOGY

All experimental work was performed with materials and electrochemical setup contained in a Vacuum Atmospheres dry box system with a helium atmosphere. Electrochemical measurements were accomplished with a Princeton Applied Research (PAR) Model 273 Potentiostat/Galvanostat interfaced with an IBM personal computer using PAR 270 Software and a Hewlett Packard Model 7550 Plus plotter. A three-electrode arrangement in a single compartment cell used a Pt foil counter electrode and Bioanalytical Systems (BAS) working electrodes of Pt, W, Au or glassy carbon. The reference electrode was an Al wire in 0.60 MEIC\AlCl₃ melt contained in a pyrex glass tube constructed with an asbestos tip to provide solution contact.

Components for the melt were synthesized and purified following the procedures developed in this laboratory (1,3). LiCl (Aldrich, 99.99+% purity) was dried for 10 days in a vacuum oven (20 inches H₂O vacuum) at 130 C. In general melts were initially prepared to be exactly neutral by adding AlCl₃ to MEIC and neutrality was verified by using cyclic voltammetry. Neutral melts were then treated with ethylaluminum dichloride (Aldrich, 97% purity) to remove protonic impurity resulting from contamination by H₂O during the synthesis and

purification of MEIC.

Buffered Neutral Melts

We expected that the composition of buffered-neutral melts might vary depending on the initial melt acidity. Buffering with LiCl starting from an acidic melt should follow the reaction

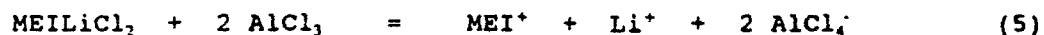


From a basic melt the reaction should be



An acidic melt (0.55 mole fraction of AlCl_3) was prepared by adding the amount of AlCl_3 calculated to give the desired composition to an exactly neutral melt which had been treated with ethylaluminum dichloride. Then LiCl was added (with a calculated 25% excess) to buffer the melt back to neutral (eq. 3). LiCl was found to dissolve much more slowly than NaCl, requiring several days with occasional heating to 50 C for attainment of a neutral melt. Neutrality of this melt was demonstrated by the absence of Al deposition and stripping during a cyclic voltammetric scan. In a slightly acidic melt Al deposition could be observed at a cathodic potential approximately one volt more negative than in 0.55 melt. This Al deposition and stripping could be observed in buffered melts even after several days (up to 5 days) of stirring with LiCl in the melt. As with NaCl, LiCl was found to be insoluble in basic melt and all attempts to buffer a melt to neutral from the basic side with LiCl were unsuccessful.

Some other possibilities for preparing a buffered-neutral melt were explored according to the following reactions:



and



Following reaction (5), equimolar amounts of MEIC and LiCl were mixed and heated. Formation of a clear colorless melt required heating to about 125 C. This melt was extremely viscous and exhibited a considerable range (approximately 50 degrees) of superheating and supercooling. On cooling the melt solidified as

a clear colorless glass at about 70 C. With repeated careful heating and cooling cycles, crystallization was induced and a white crystalline solid (assumed to be MEILiCl_2) was recovered. Crystals suitable for x-ray analysis could not be obtained from this sample. When AlCl_3 was mixed with this solid in a 2:1 molar ratio, a clear colorless melt was readily formed. Adequate proportions of the melt were not prepared for electrochemical studies but a sample was retained for NMR studies. This method of preparing a Li-containing neutral melt merits further attention.

Following the outline of reaction (6), solid LiAlCl_4 (Johnson-Matthey) was added to an exactly neutral melt in equimolar proportions. After prolonged stirring most of the solid had not dissolved in the melt. This method for preparing a Li-containing neutral melt does not appear to be fruitful.

RESULTS AND DISCUSSION

After MEIC of acceptable purity had been obtained and the proton impurity of the melt reduced to less than 5 mM by treatment with ethylaluminum dichloride, a LiCl buffered melt was prepared as described above. The buffered melt remained slightly acidic even after seven days of stirring with excess LiCl , similar to the observed behavior of NaCl buffered melts reported by Riechel and Wilkes (4). The acidity of the buffered melt was demonstrated by the deposition and stripping of Al as seen in Fig. 1, a and b. All the cyclic voltammograms presented in this report were recorded at a sweep rate of 100 mV/sec using a W working electrode (except for Fig. 5 which illustrates sweep rate dependence). Al deposition occurs at -1.32 V and stripping at -0.194 V (Fig. 1, a). The potential for Al deposition in this melt is more cathodic than the potential for the same process in more acidic melts by over one volt. The Al stripping in this melt occurred at a potential over 100 mV more cathodic than that reported for a similar NaCl buffered melt having an AlCl_3 mole fraction of 0.5002 (4), suggesting that the LiCl buffered melt has a composition between 0.5002 and exactly neutral. The deposition of Li at -1.6 V, (Fig. 1, b) significantly changes the Al stripping

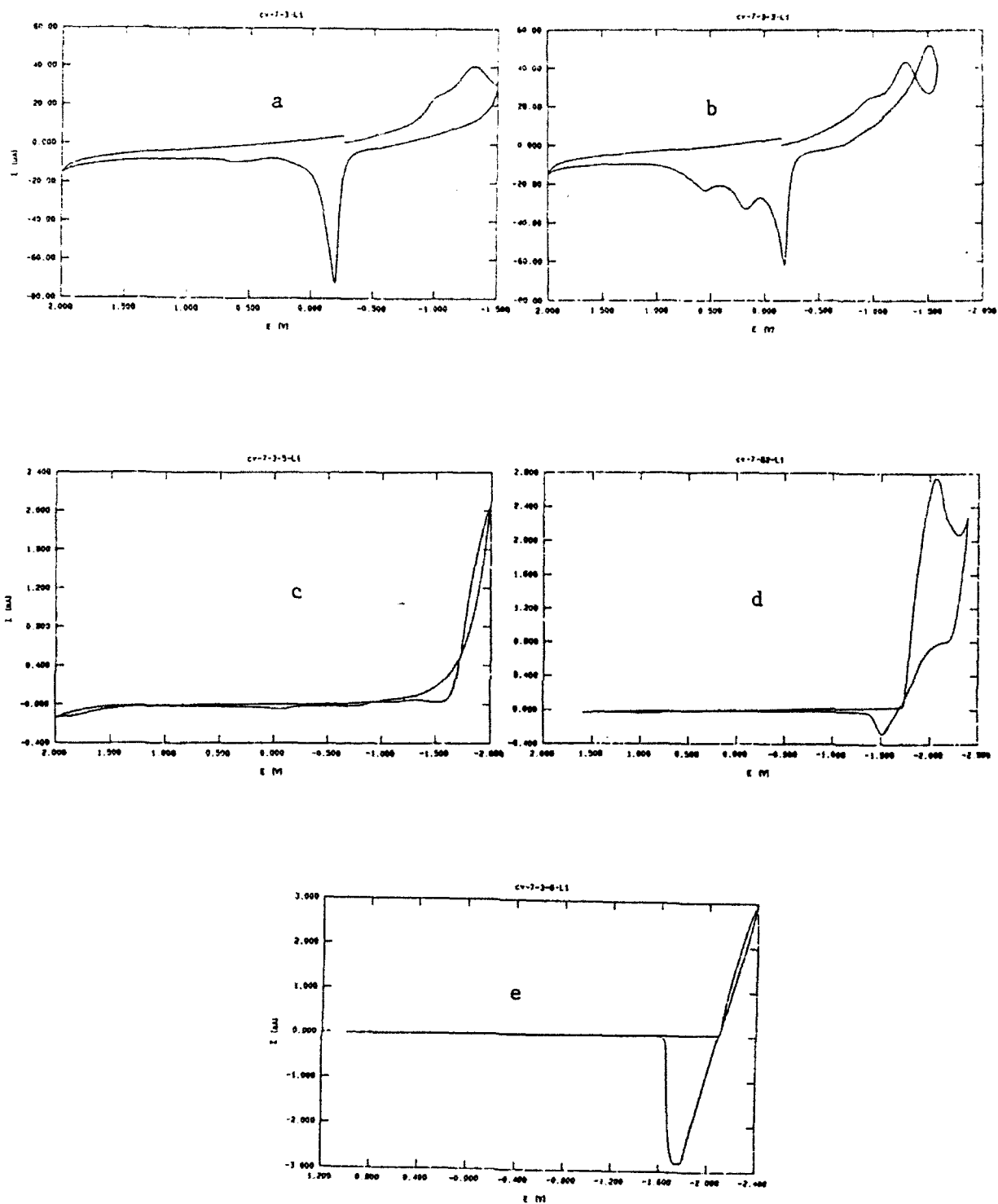


FIGURE 1

process with the appearance of two additional oxidation peaks related to Al. One might speculate that the Li reacts with Al on the electrode surface to form a Li/Al alloy which is more difficult to strip from the surface. Experiments with a Li/Al alloy would be a logical extension of this project.

Curves c-e of Fig. 1 show the effect of proton added to the buffered melt. The first noticeable effect of added proton is the loss of Al deposition and stripping, indicating that the melt became neutral. The oxidation peak at -1.55V, seen in Fig. 1,d, is not Li stripping but is the result of an oxidizable species formed by reduction of MEIC at the melt limit. In this case the melt limit obtained with a proton concentration of about 50 mM is approximately -2.35 V. When the proton concentration is sufficiently high the melt becomes milky with a very fine white solid suspended in solution. At this concentration (about 60 mM) the melt cathodic limit is significantly extended and Li stripping (Fig. 1,e) occurs typically at -1.76 V. The extension of the melt limit depends on proton concentration and was observed to be as negative as -2.80 V. Another curious effect observed consistently in melts with sufficiently high proton concentration is that proton reduction peaks typically seen at approximately -0.6 and -1.0 V at a more sensitive current setting (uA instead of mA which can be obtained by terminating the cathodic sweep before Li deposition) are no longer observed. Instead the voltammograms exhibit a linear increase of current with potential and no peaks until Li deposition is reached. Attempts to determine the precise proton concentration at which Li stripping begins failed due to relatively rapid loss of proton from the melt as HCl. Fig. 2 illustrates the change in cyclic voltammetric (CV) behavior with time as proton is lost from the melt. The data presented in Fig. 2 are representative of the changes occurring from proton loss over a one hour period, beginning with curve (a) for a melt with adequate proton concentration. The change in melt composition with time represents a major problem which must be dealt with before meaningful quantitative studies can be successfully carried out. After several additions of MEIHCl₂ to the melt the buffering capacity of the melt is exceeded since the proton leaves as HCl and the MEIC remains to make the melt basic.

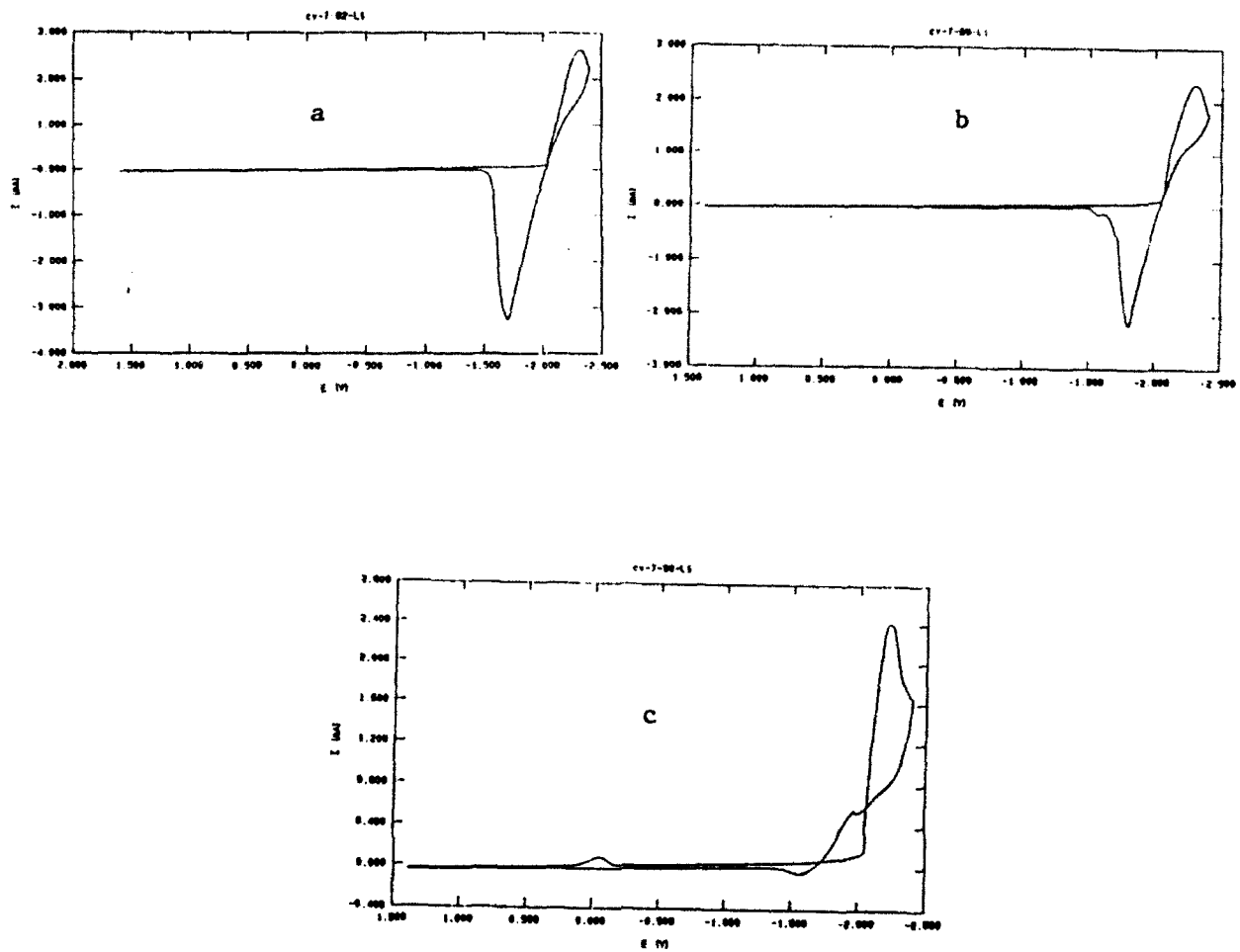


FIGURE 2

The Li deposition/stripping behavior and extended melt limit, seen in Fig. 2, a could be restored by suitable addition of AlCl_3 and LiCl to adjust the buffering capacity of the melt followed by addition of MEIHC_2 . Melts which did not have the proper composition to exhibit Li stripping and extended melt limit always resulted in a brown (copper-colored) film deposited on the electrode, even at potentials positive to the detectable melt limit. This film was easily removed by wiping the electrode but was not electrochemically removed up to the anodic melt limit. This film significantly influenced electrode behavior necessitating cleaning of the electrode after every CV to obtain reproducible results. Peak currents for Li deposition were reduced by a factor of two or more after a single CV cycle if the electrode was not cleaned by removal of this film. With proper melt composition the brown film was no longer a problem and electrode cleaning was not required. In some cases when the melt composition was intermediate between the brown film stage and that necessary for desired CV response, Li deposition resulted in coating of the electrode with a gray deposit that could not be removed by wiping. This deposit also had a significant influence on any subsequent measurements and required removal. On removal of the electrode from the dry box this gray deposit immediately disappeared when brought in contact with water. We speculate that this deposit contained Li strongly bound to the W surface, hence it could not be removed by wiping but readily dissolved in water. The reason for its electrochemical inertness is not readily apparent.

Figure 3 illustrates the CV behavior of Li deposition and stripping in a melt of proper composition as the cathodic sweep limit is increased. The cathodic potential limits in Fig. 3 are as follows: a) -2.15 V, b) -2.20 V, c) -2.30 V, d) -2.40 V, e) -2.50 V, f) -2.60 V, g) -2.70 V and h) -2.80 V. The potential for Li stripping moves less negative as the cathodic limit is increased up to -2.4 V. At cathodic limit potentials more negative than -2.4 V, the potential for Li stripping remains essentially constant at -1.7 V. In melts of the desired composition the Li stripping curve remains smooth with no breaks. The cathodic melt limit for this series is very close to -2.8 V. When the melt composition falls below a certain minimum proton concentration other variations

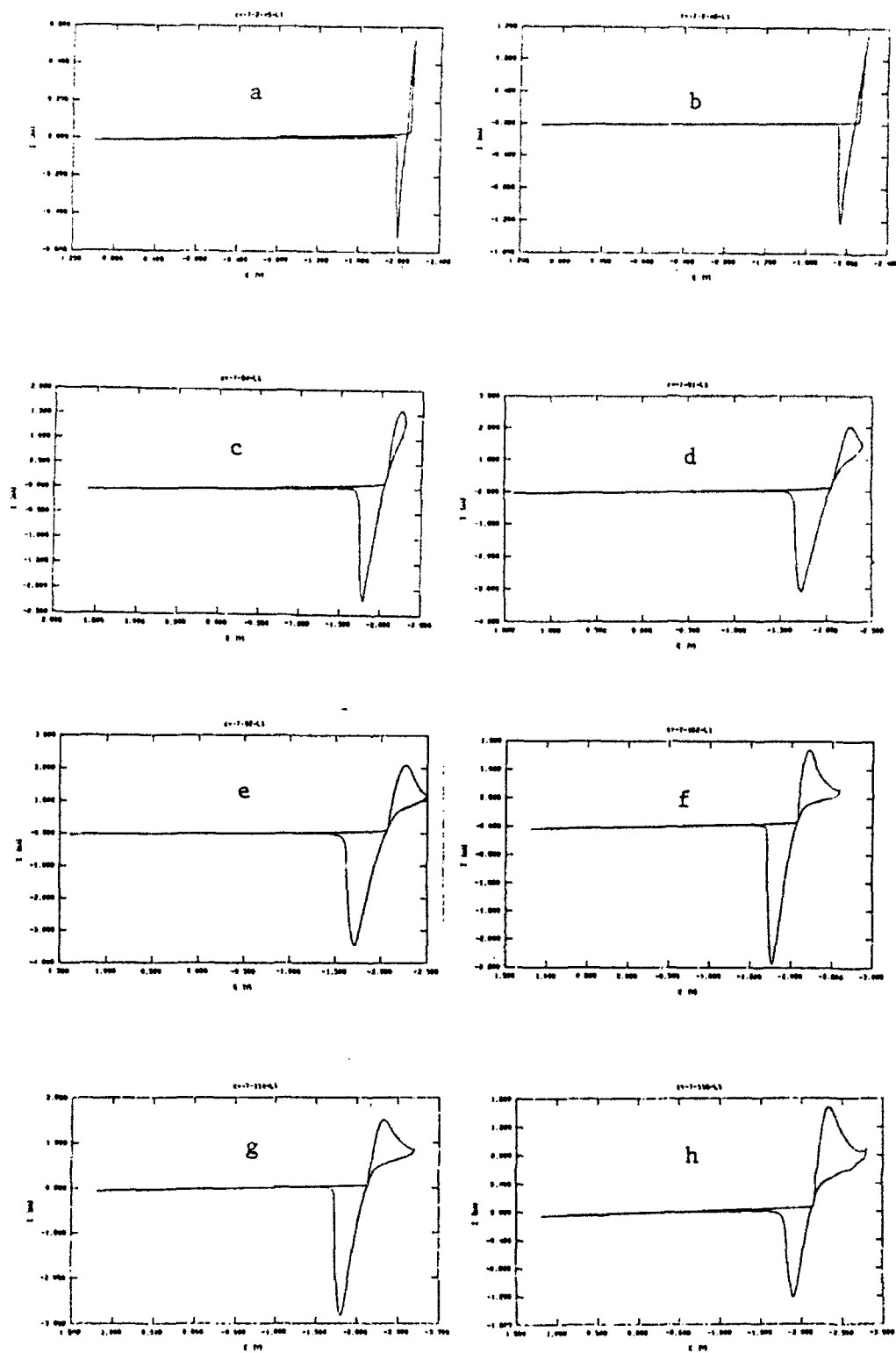


FIGURE 3

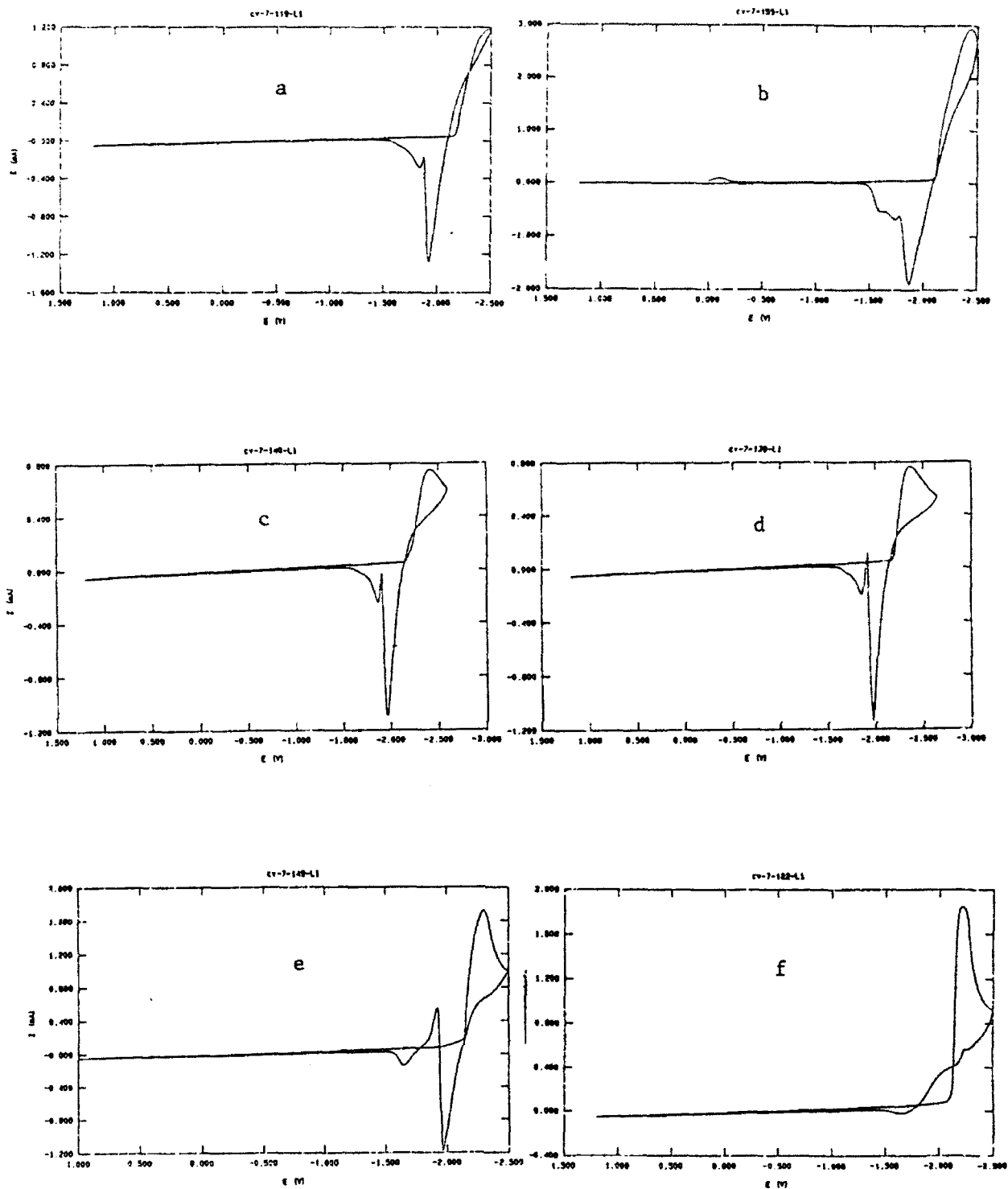


FIGURE 4

of CV behavior for Li stripping can be observed and some of these are illustrated in Fig. 4. This series from (a) to (f) shows the influence on CV curves of the loss of proton with time. It would be of interest to know why the additional oxidation peak is observed as the proton concentration decreases and why the Li stripping peak disappears. Investigation of these questions will require steps to stabilize the melt composition so that quantitative studies can be made. We are convinced that the kinds of CV curves illustrated in Fig. 4 can be prevented with proper melt composition.

The effect of sweep rate on Li deposition and stripping is illustrated in Fig. 5. For each CV in this series the cathodic limit was -2.50 V and the sweep rates were changed as follows: (a) 10 mV/sec, (b) 20 mV/sec, (c) 50 mV/sec, (d) 100 mV/sec, (e) 250 mV/sec and (f) 500 mV/sec. The unusual stripping behavior obtained at 10 mV/sec was reproducible and appears to be real. Similar results were obtained at sweep rates of 5 mV/sec and 2 mV/sec. CV behavior at sweep rates greater than 500 mV/sec were qualitatively the same as for 500. Quantitative data for sweep rate dependence have been collected in Table I.

TABLE I

Sweep Rate Dependence

mv/sec	E_{begin} (V)	Li Deposition			Li Stripping		
		E_{peak} (V)	I_{peak} (mA)	A_{peak} (mC)	E_{peak} (V)	I_{peak} (mA)	A_{peak} (mC)
5	-2.09	-2.23	0.85	91.3	-1.94	0.80	17.4
10	-2.10	-2.34	1.13	53.9	-1.90	1.20	18.4
20	-2.10	-2.34	1.51	34.6	-1.82	1.88	17.9
50	-2.11	-2.35	2.07	18.6	-1.80	2.77	12.7
100	-2.12	-2.44	2.67	12.8	-1.76	3.16	9.00
250	-2.14	*	3.53	6.21	-1.67	4.01	4.71
500	-2.16	*	3.67	3.36	-1.67	4.14	2.45
1000	-2.16	*	3.52	1.55	-1.68	4.13	1.16

* Cathodic limit was set at -2.50V to avoid reduction of melt.

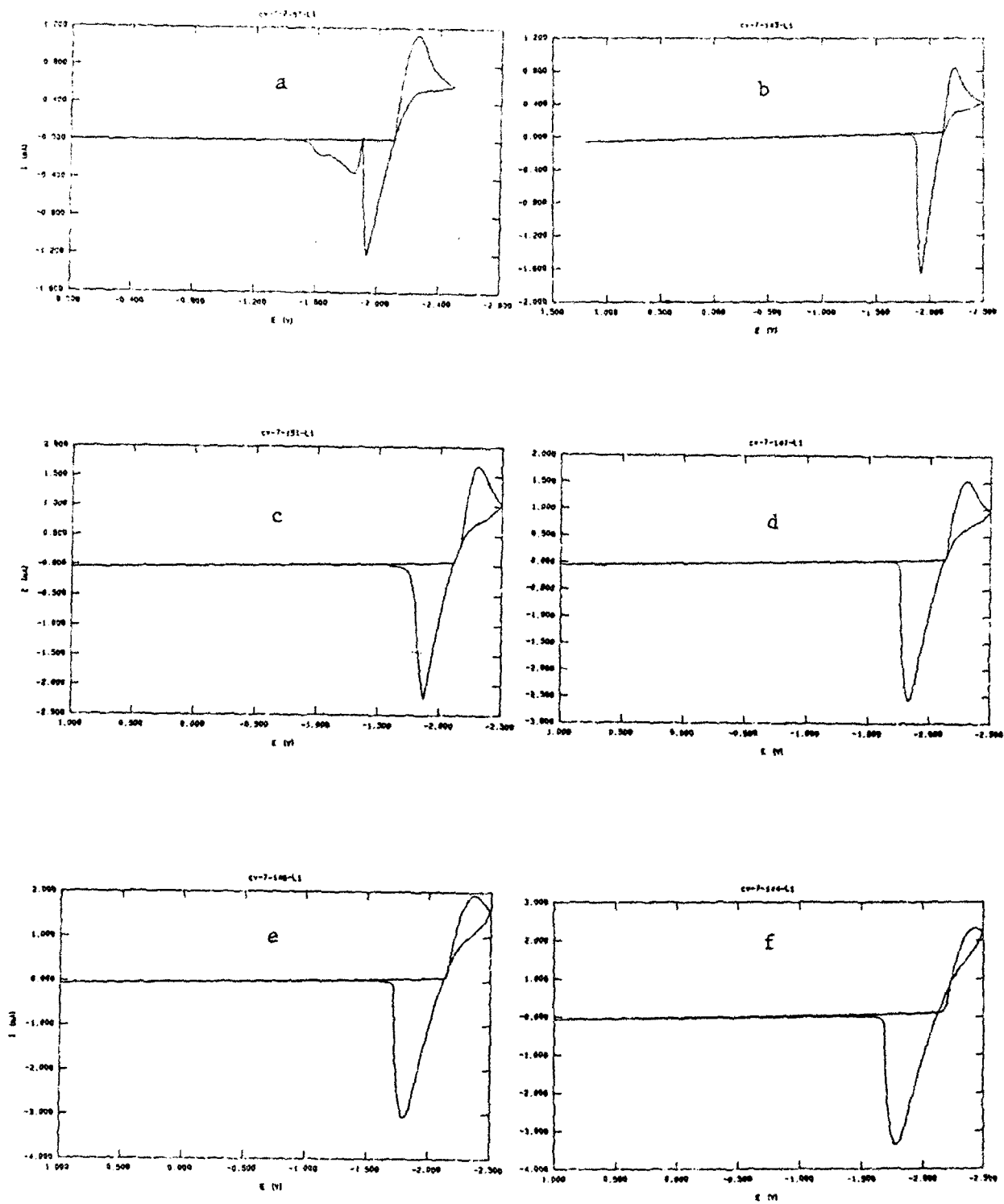


FIGURE 5

The potential at which Li deposition begins shifts slightly more cathodic with increasing sweep rate. The difference in peak potential for deposition and stripping shows a marked dependence increasing from 290 mV at 5 mV/sec to 680 mV at 100 mV/sec. The variation of deposition peak current with square root of the sweep rate is constant within experimental error, indicating that Li deposition is diffusion controlled. These observations are consistent with those reported by Carlin (5). The charge ratio for stripping/deposition increases from about 0.20 at 5 mV/sec to a more or less constant value of about 0.75 at sweep rates greater than 100 mV/sec. The charge ratio is also influenced by the cathodic limit of the CV and this is summarized in Table II. The efficiency for stripping appears to reach a maximum at a sweep limit of -2.50 V.

TABLE II

Dependence Of $Q_{\text{stripping}}/Q_{\text{deposition}}$
from Cyclic Voltammetry on Cathodic Sweep Limit

Cathodic Limit (V)	$Q_{\text{stripping}}/Q_{\text{deposition}}$
-2.20	0.33
-2.30	0.66
-2.40	0.70
-2.50	0.73
-2.60	0.59
-2.70	0.60
-2.80	0.22

Charge Ratios are valid to ± 0.05

Double pulse chronoamperometry was used to further study the stability of deposited lithium. Fig. 6 illustrates typical results for Li stripping at -1.8V after deposition for 3 sec at: (a) -2.15 V, (b) -2.175 V, (c) -2.30 V and (d) -2.40 V. The results summarized in Table III show that the efficiency of Li stripping is greatest following deposition at -2.25 V. We also see that the Li deposit is highly unstable as demonstrated by the marked effect of a time delay between deposition and stripping.

TABLE III

Double Pulse Chronoamperometry

Dependence of $Q_{\text{stripping}}/Q_{\text{deposition}}$ on deposition potential and delay between deposition and stripping.

Deposition Potential	$\frac{Q_{\text{stripping}}}{Q_{\text{deposition}}}$	Time Between Pulses at Open Circuit
-2.15 V	0.40	0 sec
-2.175	0.65	"
-2.20	0.73	"
-2.25	0.76	"
-2.30	0.73	"
-2.40	0.70	"
-2.50	0.56	"
-2.60	0.41	"
-2.20	0.75	10 sec
"	0.62	30 "
"	0.37	60 "
"	0.12	300 "

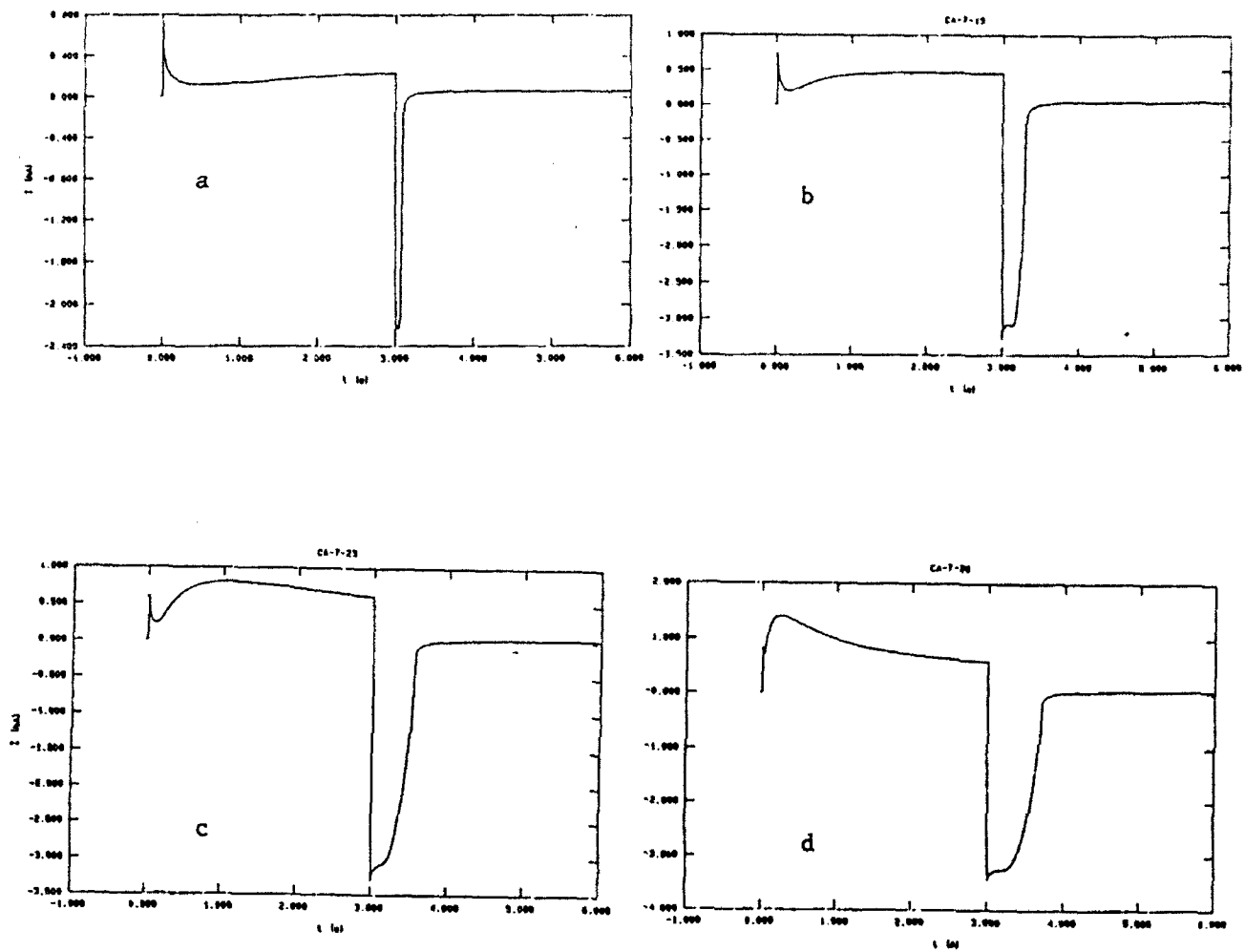


FIGURE 6

The instability of the deposited lithium in the buffered neutral melt is a troubling problem that must be solved before Li can be considered as a viable battery anode. Li deposits equivalent to 30-50 monolayers (calculated from electrode area, deposition current and time) appear to be lost from the electrode surface in less than 5 minutes. The open circuit potential established at a W electrode after Li is deposited maintains a fairly constant potential of approximately -2.00 V for only 60-90 sec. After 90 sec. the open circuit potential falls rapidly to -.35 to -.40 V, the value established at a clean W electrode. This change in open circuit potential is not dependent on the deposition potential over the range of -2.2 to -2.8 V, or on deposition times from 1 sec. to 10 sec., suggesting that it is not dependent on the quantity of Li deposited. The melt composition which yields good deposition/stripping behavior for Li does not appear to provide equally acceptable conditions for Li stability. The data obtained on the stability of deposited Li in buffered neutral melt is substantially different than of deposited Na. Sodium deposited on W from NaCl buffered melt is stable for at least 5-6 hours (2).

In an initial attempt to improve melt stability Li_2CO_3 equivalent to the calculated quantity of proton was added to the buffered melt. The expected reactions are



with the Cl^- buffered by Li^+



This procedure proved very effective for complexing the proton in the melt, but the $\text{Al}_2\text{Cl}_7\text{OH}$ did not release the proton as we had hoped. Within 30 min. after addition of Li_2CO_3 , no Li stripping could be obtained. The development of a stable proton source for the buffered melt is required, not only for a Li anode, but also for other alkali metal anodes.

This discussion is concluded with a few brief comments on the CV behavior of Pt, Au and glassy carbon electrodes. Pt electrodes gave results qualitatively identical to those presented here for W. Li deposition began at about the same potential and stripping curves were identical to those recorded on W. Anomalous

behavior similiar to that shown for W (Fig. 4) was also obtained on Pt when the melt composition was not properly maintained. No identifiable deposition or stripping of Li could be obtained with glassy carbon. CV behavior at a Au electrode showed Li deposition beginning at -2.00 V, more than 100 mV more positive than for Pt or W. Li stripping from Au was significantly different than for Pt or W and was not reproducible. The primary feature of stripping from Au was a primary peak at -1.45 V with shoulders at -1.65 V and -1.81 V. These three stripping peaks were always present but their relative heights varied over a wide range. Time limitations did not permit further work with these electrodes.

Conclusions

1. The primary conclusion from this study is that Li deposited on W or Pt electrodes is not stable under any of the conditions used for the LiCl buffered-neutral melts. The highest efficiencies for stripping Li deposited on W were in the area of 75%. Since no detectable traces of Li were left on the electrode we conclude that it was lost by another process, most likely reaction with some component in the melt.

2. In most respects, the behavior of LiCl buffered-neutral melts is very similiar to that of NaCl buffered-neutral melts. The one significant difference is the much greater stability of deposited sodium. The problem of Li instability in buffered neutral melt must be solved before Li can be considered a viable anode material.

3. The addition of proton to the melt is a key factor in obtaining quasi-reversible deposition and stripping of Li. The rapid loss of proton from the melt as HCl is another significant problem that must be solved before alkali metals can be used as reversible anodes in these melts.

4. The brown film formed on W by cathodic reduction of the melt when the melt does not have the proper composition to give Li stripping or extension of the melt limit is not seen when the melt composition is correctly maintained. While the exact composition of the melt termed "proper composition" is not easily specified or easily maintained, it can be achieved with relative ease.

References

1. Melton, T.J., Joyce, J., Maloy, J.T., Boon, J.A. and Wilkes, J.S., J. Electrochem. Soc. 137, 3865 (1990)
2. Riechel, T.L. and Wilkes, J.S., J. Electrochem. Soc. 139, 977 (1992)
3. Wilkes, J.S., Levisky, J.A., Wilson, R.A. and Hussey, C.L., Inorg. Chem. 21, 1263 (1982)
4. Riechel, T.L. and Wilkes, J.S., unpublished manuscript
5. Scordilis-Kelley, C., Fuller, J., Carlin, R.T. and Wilkes, J.S., J. Electrochem. Soc. 139, 694 (1992)

WILFORD HALL MEDICAL CENTER

ENHANCED PHYSIOLOGIC MONITORING OF CLOSED HEAD-INJURY

**Michael L. Daley
Associate Professor
and
Brian Carter
Graduate Student
Department of Electrical Engineering**

**Memphis State University
Memphis, TN 38152**

**Final Report for:
Summer Research Program
Wilford Hall Medical Center
Lackland AFB, TX**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.**

September, 1992

ENHANCED PHYSIOLOGIC MONITORING OF CLOSED HEAD-INJURY

Michael L. Daley
Associate Professor
and
Brian Carter
Graduate Student
Department of Electrical Engineering
Memphis State University

Abstract

The aim of this study was to develop both a laboratory model of closed head-injury and an analytical model of venous blood flow from the brain to test the *hypothesis that variations in venous pressure associated with the respiratory cycle can have a dominant influence on venous flow from the brain during elevated intracranial pressure*. A young adult pig with an implanted intracranial balloon designed to manipulate intracranial volume was used as a laboratory model. An analog electrical circuit model was used to provide a theoretical analytical description of cerebral venous blood flow during elevated intracranial pressure. Both experimental and theoretical results indicate that during intact autoregulation of cerebral blood flow, respiratory induced venous pressure changes systematically influence intracranial blood volume. Specifically, intracranial blood volume increases during inhalation and decreases during expiration. Furthermore, the difference in change of intracranial volume between the two phases of ventilation, inhalation and expiration, increases with increasing mean intracranial pressure. However, during loss of regulation of cerebral blood flow, venous blood flow and the resulting changes of intracranial blood volume are not systematically influenced by respiratory induced venous pressure changes.

ENHANCED PHYSIOLOGIC MONITORING OF CLOSED HEAD-INJURY

Michael L. Daley

INTRODUCTION:

During the routine neurosurgical intensive care of the patient with head-injury accurate monitoring of physiologic variables is accomplished by specially trained staff with the use of precision bedside instrumentation. The general aim of this study is to develop techniques which characterize venous blood flow from the brain during elevated intracranial pressure from an analysis of the physiologic signals provided by the bedside monitor. Such techniques would be of value in the recognition of the loss of regulation of cerebral blood flow. The specific aim of this study is to use both a laboratory model of the clinical condition and a laminar flow model of venous blood flow from the brain to test the hypothesis that during autoregulation of cerebral blood flow, variations in venous pressure associated with the respiratory cycle systematically influences venous flow from the brain during elevated intracranial pressure.

The return of blood from the mammalian brain generally travels the following path through the venous vasculature to the heart: cerebral veins to leptomeningeal (bridging) veins to cranial venous sinuses to internal jugular vein to superior vena cava to heart. Because blood is relatively viscous and under low pressure in the venous vasculature, laminar flow is assumed. As a result an electrical analog circuit model of the dynamics of blood flow through the brain has been proposed [1]. A modification of this proposed model which solely describes the dynamics of venous blood from the brain is shown in Fig. 1. The voltage generator, $P_{cv}(t)$, represents cerebral venous pressure. During elevated intracranial pressure, the leptomeningeal (bridging) veins undergo a selective "cuff constriction" which increases pressure in the cerebral veins and prevents them from collapsing [2,3]. As a result, the cerebral venous pressure is approximately equal to intracranial pressure. The resistance, $R_{\text{leptomeningeal}}$, represents the variable pressure dependent resistance of the leptomeningeal veins. The capacitance, $C_{\text{venous compliance}}$, represents the compliance of the venous system from the cranial sinuses to the heart. The charge on this capacitor represents volume of venous blood. The resistor, R_{external} , represents the external venous resistance from the sinuses to the heart. Finally, the voltage generator, $P_{\text{resp}}(t)$, represents the change in venous pressure associated with respiration. Under conditions of minimally elevated intracranial pressure, "cuff constriction"

MODEL OF CEREBRAL VENOUS BLOOD FLOW

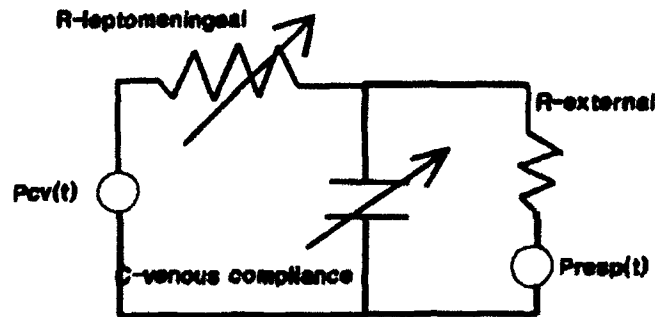


Figure 1. A Model for Venous Hemodynamics During Elevated Intracranial Pressure. The resistance, $R_{\text{leptomeningeal}}$, represents the resistance of the bridging veins between the cerebral veins and cranial sinuses. Because these veins selectively compress during elevated intracranial pressure, this resistance is dependent on intracranial pressure. Also, the capacitance, $C_{\text{venous compliance}}$, represents the compliance of the venous system from the cerebral veins to the heart. This compliance is dependent on partial pressure of blood carbon dioxide.

by the bridging veins is minimal. However, as intracranial pressure increases, "cuff constriction" by the bridging veins results in reduced cranial sinus pressure, and changes in venous pressure associated with respiration begin to influence venous flow. In particular, the model predicts that during inhalation cerebral blood volume increases and during expiration cerebral blood volume decreases. Furthermore, as intracranial pressure increases, the difference in cerebral blood volume change between the two phases of respiration will increase. The purpose of this study was to experimentally test the validity of the proposed model.

METHODS AND MATERIALS:

Animal Model of Closed Head Injury :

Nine adult pigs ranging in weight from 15 to 40 kg were used in this study. The procedures used were similar to those described for other animal studies employing this model [4-7]. Animals were anesthetized, intubated, paralyzed, and placed on a ventilator. A pressure recording via a fluid-filled catheter inserted in the brachial artery and a three lead electrocardiographic signal were used to record arterial pressure and ECG activity. In addition, a catheter was also placed in the internal jugular vein to record venous pressure. To control intracranial pressure, the skull was trephined and a small inflatable balloon was extradurally placed. Intracranial pressure was recorded using the Camino fiber optics system.

Recording Techniques:

An IBM compatible 486, 33 MHZ, with four analog to digital channels was connected to an EVR Physiologic Monitor. The four physiologic signals, ECG, arterial pressure, intracranial pressure, and jugular venous pressure, were simultaneously acquired at a rate of 250 samples/sec.

Analysis of Physiological Signals:

The procedures used for digitizing and screening out artifacts, which were minimal because of the experimental setting, were similar to those previously described [8,9]. A recently developed method for deriving an estimate of pulsatile cerebral blood volume change from the intracranial pressure signal was used [10]. The theoretical basis of this method is based on the assumption that the pressure volume characteristic of the craniospinal sac can be reasonably approximated by a closed-volume system with a pressure volume relationship described as:

$$P(V(t)) = P_0 \exp(E(V(t) - V_{eq})) \quad (1).$$

The derivative with respect to time of this relationship is

$$dP(V(t))/dt = E \cdot P_0 \exp(E(V(t) - V_{eq})) \cdot dV(t)/dt \quad (2).$$

By dividing equation (1) into equation (2) and rearranging the terms, an expression for the time derivative of volume is obtained:

$$dV(t)/dt = (1/E) \cdot (dP(V(t))/dt) / P(V(t)) \quad (3).$$

Using standard numerical methods, the above expression is determined from the digitized clinical ICP recording. Integrating the above expression, eqn. 3, with respect to time, yields a signal proportional to volume. One series of published measures of the elastance constant for adults indicates an average value of about 0.25/ml [11]. The validity of the assumption that an exponential relationship exists between intracranial pressure and intracranial volume can be tested. In particular, the pulsatile blood volume change over each cardiac cycle represents a small volume perturbation relative to the overall volume contained in the craniospinal sac. Thus the exponential relationship can be approximated by a first-order polynomial. Therefore, changes in the contour of the volume signal should correspond to contour changes in the intracranial pressure signal. Recordings of the ECG, the arterial and intracranial pressure signals and the corresponding intracranial volume signal are shown in Figure 2. The similarity in the contour of the two signals is evident.

ECG, Arterial, ICP, and Cerebral Blood Vol. Signals.

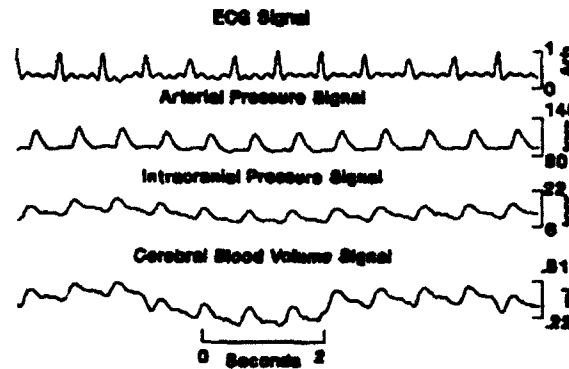


Figure 2. Electrocardiographic (ECG), Arterial, ICP, and Cerebral Blood Volume Signals. The pulsation of intracranial pressure associated with each heart beat reflects the brief net change of cerebral blood volume caused by pulsatile arterial flow into and venous flow out of the craniospinal sac during each cardiac cycle. This transient change in cerebral blood volume represents a small perturbation relative to the overall volume contained in the craniospinal sac. For this condition the experimental relationship between intracranial pressure and volume can be approximated by a first-order polynomial. Thus the similarities of the contours of the intracranial pressure and cerebral blood volume signal are expected.

RESULTS:

Volume Calculations:

Using the R-waves as time markers of the cardiac cycle the change of intracranial volume over one cardiac cycle can be computed (see Fig. 3). Because of the fluctuations in venous pressure these estimates of pulsatile volume change over each cardiac cycle would vary with respiration as shown. For

ECG, Cerebral Blood Volume, Mean ICP, and Pulsation Volume

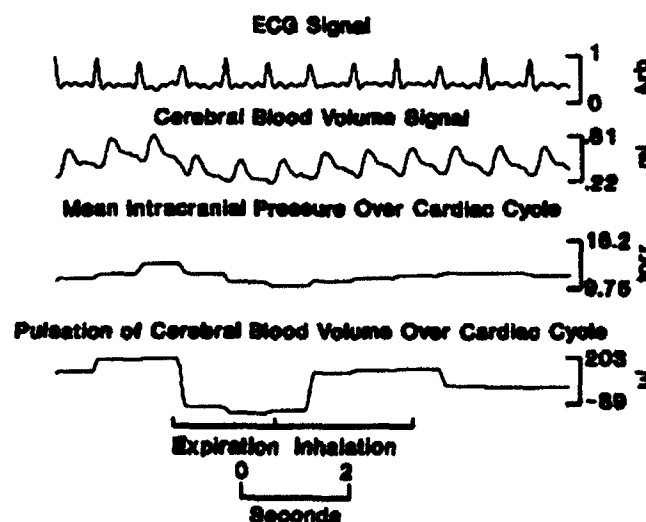


Figure 3. Electrocardiographic (ECG), Cerebral Blood Volume, Mean Intracranial Pressure, and Magnitude of the Pulsation Volume Signals. During intact regulation of cerebral blood flow, cyclic changes in venous pressure over the respiratory cycle influence venous flow. Because of the variation in venous pressure both mean intracranial pressure over each cardiac cycle and the magnitude of the pulsation of cerebral blood volume vary during ventilation. Both measures reach a minimum during expiration and a maximum during inhalation.

this example, during the inhalation phase, pulsatile change in cerebral blood volume over each cardiac cycle reached a maximum of approximately 200 μ l. In contrast during the expiration phase, change in cerebral blood over each cardiac cycle reached a minimum of approximately -90 μ l over a cardiac cycle. Also note that peak intracranial pressure increased during inhalation and decreased during expiration. To calculate change of volume during each phase of the respiratory cycle, characteristics of the jugular pressure signal were used to detect the beginning and end of inhalation and the end of the expiration phase. In addition, knowledge of the respiration rate and I:E ratio was used. Start of the inhalation phase was denoted by the location of a minimum in the jugular trace. The end of the inhalation phase was denoted by the occurrence of a maximum at a point in time expected from knowledge of respiration rate and the I:E ratio. The expiration phase of ventilation was marked in time by the end of the inhalation phase and the occurrence of a minimum at a point in time expected from knowledge of the rate of respiration. Using this technique, changes in volume for each phase of respiration were calculated from the physiologic signals recorded during intact regulation of cerebral blood flow and periods of steady-state intracranial pressure (see Figure 4). As expected from the hypothesis of this study, difference in intracranial volume between the two phases of respiration was found to increase with increasing mean intracranial pressure.

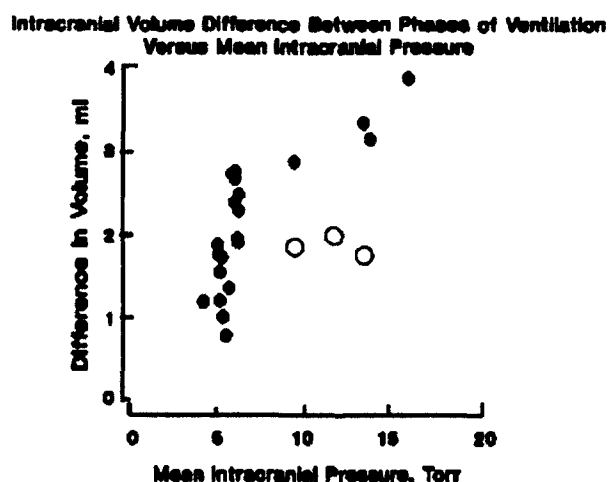


Figure 4. Intracranial Volume Difference between the Phases of Ventilation Versus Mean Intracranial Pressure. During intact regulation of cerebral blood flow, changes of intracranial volume over the respiratory cycle increase with increasing mean intracranial pressure (filled circles). During deregulation of cerebral blood flow, systematic variation of intracranial volume with increasing mean intracranial pressure was not evident (open circles).

However, following deregulation of cerebral blood flow, intracranial volume changes over the respiratory cycle were not systematically influenced by increasing mean intracranial pressure.

Recognition of Deregulation of Cerebral Blood Flow:

In these experiments, the recognition of deregulation of cerebral blood flow was based on the presence of salient pulsations associated with each cardiac cycle in the jugular venous pressure recordings. These pulsations occurred after the intracranial pressure was maintained above arterial pressure for a brief period. A comparison of recordings of intracranial pressure and jugular venous pressure before and after manipulation of intracranial volume and during normocapnia is presented in Figure 5. In some cases,

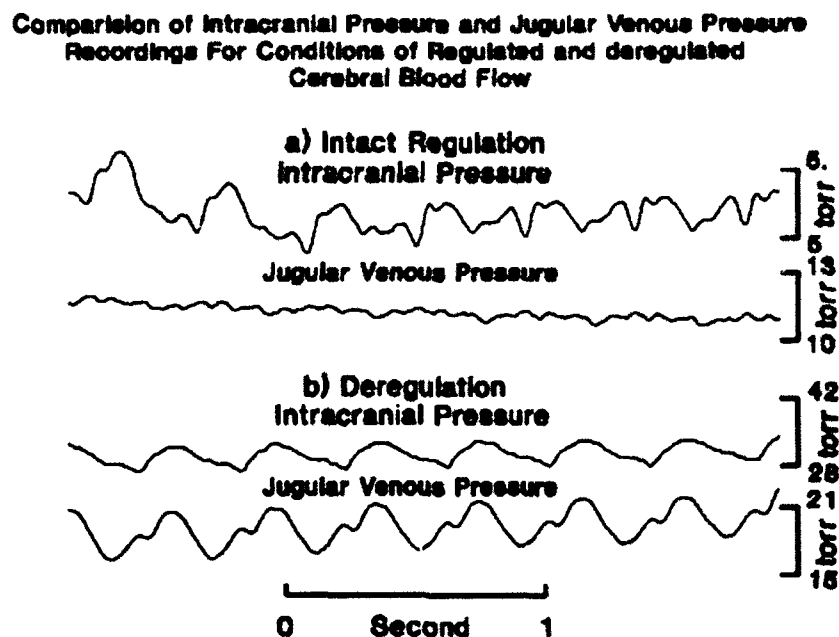


Figure 5. Comparison of Intracranial Pressure and Jugular Venous Pressure Recordings for Conditions of Regulated and Deregulated Cerebral Blood Flow. a) Intact Regulation of Cerebral Blood Flow: During regulation and normocapnia, pulsations in jugular pressure are minimal. b) Deregulation of Cerebral Blood Flow: This condition is marked by salient pulsations of jugular venous pressure and produced by maintaining intracranial pressure above arterial pressure for a period of one to two minutes.

after deregulation occurred, the salient pulsations of jugular pressure sometimes disappeared. During and immediately following the deregulation condition, respiratory induced venous pressure changes did not significantly influence intracranial blood volume (see Figure 4). However, changes in intracranial pressure and intracranial blood volume correlated with changes in arterial pressure. To further illustrate the deregulated condition, a 10 minute period during which intracranial volume was manipulated by infusion into the intracranially placed balloon was selected for analysis. For this time interval average total volume,

arterial pressure, jugular venous pressure, and intracranial pressure were computed for each respiratory cycle of approximately five seconds. This example illustrates the strong positive correlation between arterial pressure and intracranial volume and the lack of a correlation between jugular venous pressure and intracranial volume during a period of deregulation (see Figure 6). Prior to the 10 minute recording

Average Total Volume, Arterial, Jugular, and Intracranial Pressure Over Each Respiratory Cycle

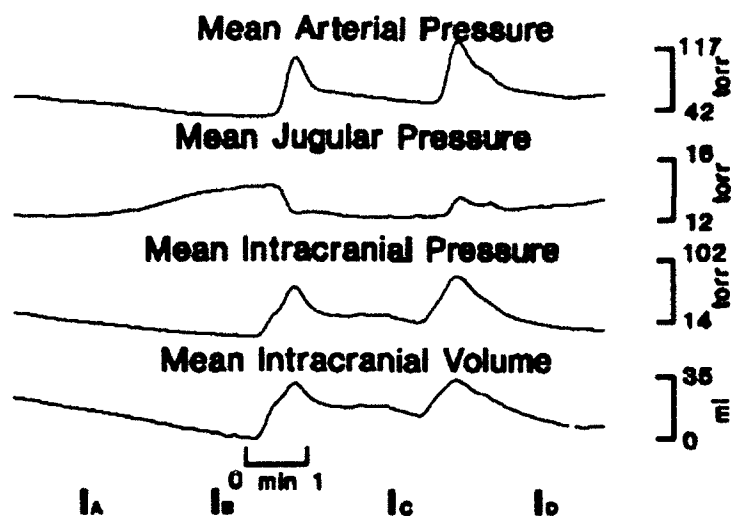


Figure 6. Average Total Volume, Arterial Pressure, Jugular Venous Pressure, and Intracranial Pressure Over Each Respiratory Cycle During a Ten Minute Period. Following points B and C in time intracranial volume was manipulated by infusion of an aliquot of fluid into the intracranial balloon. Mean arterial pressure positively correlated with changes of intracranial volume. In contrast, mean jugular venous pressure did not correlate with changes of intracranial volume.

period, intracranial pressure was elevated and briefly maintained above intracranial pressure, and deregulation occurred as denoted by the appearance of salient pulsations in the jugular venous pressure recording. During the first time interval denoted by points A and B, the inflated intracranial balloon was held at a constant volume and arterial pressure, intracranial volume, and intracranial pressure decreased while jugular venous pressure increased. Just following point in time denoted by B approximately 8 ml of fluid was rapidly infused into the intracranial balloon. As expected intracranial volume and intracranial pressure increased dramatically. After a slight delay marked by an inflection point occurring in both the intracranial volume and pressure traces which reflected the time for intracranial pressure to rise above arterial pressure, jugular venous pressure rapidly decreased to a plateau and simultaneously arterial pressure sharply increased to a peak of over 100 Torr coincident with the peak of approximately 30 ml of intracranial volume. The second infusion of the intracranial balloon occurred after point C in time. The rise

of intracranial volume and pressure was accompanied by a sharp rise of arterial peak with all three signals reaching a peak simultaneously. In contrast jugular venous pressure rose moderately to a plateau.

In summary, during and immediately following deregulation of cerebral blood flow, changes of arterial pressure positively correlated with changes of intracranial volume. In contrast, over this same time period, a correlation between jugular venous pressure changes and arterial pressure changes was not evident.

Model of Cerebral Blood Flow :

Preliminary work using the analytical model (see Figure 1) and experimental data has also been accomplished. Specifically, a fourth-order Runge-Kutta method was used to solve the following system equation describing the dynamics of cranial sinus pressure:

$$\frac{dP(t)_{vs}}{dt} = -\frac{(R_{lep} + R_{ext})}{R_{lep} * R_{ext}} P(t)_{vs} + \frac{P(t)_{cv}}{C_v * R_{lep}} + \frac{P(t)_{resp}}{C_v * R_{ext}} \quad (4).$$

Since the leptomeningeal network throughout the brain and the compliance of the venous system is distributed, the corresponding resistance and capacitance of these model components were randomly perturbed about a mean value. In this model, cerebral venous pressure, $P_{cv}(t)$, is assumed equal to the intracranial pressure signal, and the variation of venous pressure over the respiratory cycle, $P_{resp}(t)$, is estimated from the jugular venous pressure recording. An example of the predicted cranial sinus pressure, which because of the assumption of laminar flow is directly proportional to venous flow, is shown here.

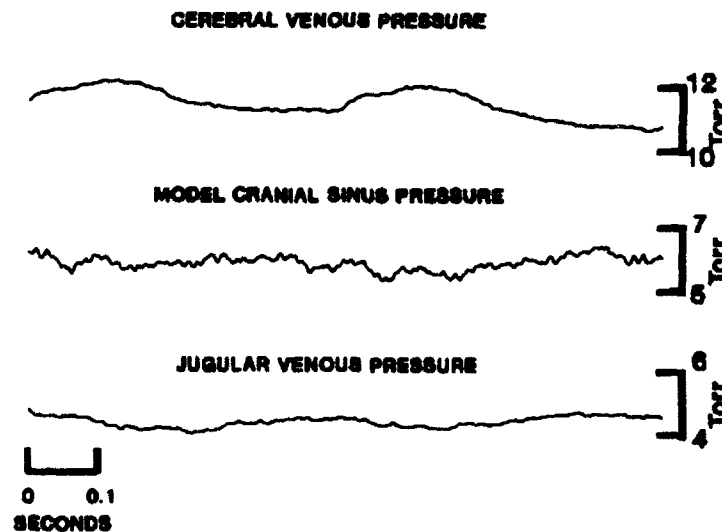


Figure 7. Preliminary Predictions of Model of Venous Hemodynamics During Elevated Intracranial Pressure. Predicted cranial sinus pressure is directly proportional to venous flow. For the condition of high values of leptomeningeal resistance, venous flow and the corresponding changes of venous blood volume are systematically influenced by changes of venous pressure associated with respiration.

DISCUSSION:

These preliminary data were obtained from an animal model in which intracranial volume was manipulated by infusion and withdrawal of fluid from an extradural balloon. For the condition in which intracranial pressure was minimally elevated without previous elevation, changes of intracranial volume between the two phases of the respiratory cycle were systematically influenced by increasing intracranial pressure. Specifically, for this condition, autoregulation of cerebral blood flow was assumed to be intact. Intracranial blood volume was found to increase during inhalation and decrease during expiration corresponding to the variations of jugular pressure occurring over the respiratory cycle. Furthermore, in support of the hypothesis of this study, the difference in change of intracranial volume between the two phases of ventilation, inhalation and expiration, was found to increase with increasing mean intracranial pressure.

After intracranial pressure was elevated above arterial pressure for more than one to two minutes and then lowered, generally venous pressure became elevated and pulsations of jugular pressure corresponding to pulsations of intracranial pressure were evident. In addition, changes of intracranial pressure followed changes of arterial pressure and not jugular venous pressure. The gain between intracranial pressure and arterial pressure was also high. This physiologic condition was assumed to represent loss of cerebral regulation of blood flow. Preliminary results indicate that during loss of regulation of cerebral blood flow, venous blood flow and the resulting changes of intracranial blood volume were not systematically influenced by respiratory induced venous pressure changes. These findings are consistent with partial or complete loss of autoregulation.

Further analysis of the physiologic recordings acquired during this study will enable a more thorough test of the hypothesis and the development of detailed parameters of the model. The preliminary predictions of the model of cranial sinus pressure appear to be reasonable. The focus of future work in this area will be to: 1) characterize the dynamics of compression and dilation of the leptomeningeal network; 2) define a relationship between venous compliance and the partial pressure of CO_2 in the arterial blood; and 3) further develop the method of approximating respiratory-induced venous pressure modulations from knowledge of intracranial pressure variation, respiration rate, and I:E ratio.

ACKNOWLEDGMENTS:

We wish to thank Dr. Michael Griffith, co-principal investigator, Department of Neurosurgery, Wilford Hall Medical Center, Col. John H. Cissik, Director, Clinical Investigations, and Sergeant J. Phelps and Edith Torres of Clinical Investigations for their support throughout the course of this work.

REFERENCES:

1. Ursino, M.: A mathematical study of human intracranial hydrodynamics Part I-The cerebrospinal fluid pulse pressure. Annals of Biomedical Engineering 16: 379-401, 1988.
2. Nakagawa, Y., Tsuru, M., and Yada, K.: Site and mechanism for compression of the venous system during experimental intracranial hypertension. J. Neurosurgery 41: 427-434, 1974.
3. Laas, R. and Arnold, H.: Compression of the outlets of the leptomeningeal veins-the cause of intracranial plateau wave. Acta Neurochir 58: 187, 1981.
4. Avezatt C.J.J. and van Eijndhoven J.H.M.: Cerebrospinal fluid pressure and intracranial volume-pressure relationships. J. Neurology, Neurosurgery and Psychiatry 42: 687-700, 1979.
5. Van Eijndhoven, J.H. and Avezatt, C.J.J.: Cerebrospinal fluid pulse pressure and the pulsatile variation in cerebral blood volume: An experimental study in dogs. Neurosurgery 19: 507-522, 1986.
6. Bircher, N.G. and Safar, P.: Intracranial pressure and other variables during simultaneous ventilation-compression cardiopulmonary resuscitation. In Miller, J. D., Teadale, G.M., Rowan, J., Gailbraith, S.L., and Mendelow, A.D. (eds) Intracranial Pressure VI, Berlin, Springer-Verlag, 1986, pp. 747-749.
7. Schedl, R., Baum, M., Benzer, H., Fasol, P., Ittner, G., Mutz, N., and Spangler, H.: Elimination of Ventilator-Related ICP-Fluctuations by Special Techniques of Artificial Ventilation. In Miller, J. D., Teadale, G.M., Rowan, J.O., Gailbraith, S.L., and Mendelow, A.D. (eds) Intracranial Pressure VI, Berlin, Springer-Verlag, 1986, pp. 747-749.
8. Daley, M.L., Gallo, A.E., Gehling, G.F., Isom, J.B., Mauch, W., and Kingsley, P.D.: Fluctuation of intracranial pressure associated with the cardiac cycle. Neurosurgery 11: 617-621, 1982.
9. Daley, M.L., Gallo, A., and Mauch, W.: Analysis of the intracranial pressure pulsation associated with the cardiac cycle. Innovation Et Technologie En Biologie Et Medicine 7: 537-544, 1986.
10. Daley, M. L. and Robertson, J.T.: Estimate of pulsatile cerebral blood volume change derived from the intracranial pressure signal. IEEE Engr. in Medicine and Biology Conference pp. 1649-1651, Oct. 1991.
11. Avezatt C.J.J. and van Eijndhoven J.H.M.: Clinical observations on the relationship between cerebrospinal fluid pulse pressure and intracranial pressure (Chap. 4). In Cerebrospinal Fluid Pulse Pressure and Cranio-spinal Dynamics: A Theoretical, Clinical and Experimental Study. Thesis, Erasmus University, Rotterdam, 1984.

**TEMPERATURE EFFECTS ON AQ. POLYMER AND BIOPOLYMER
SOLUTION VISCOSITIES AND ON ERYTHROCYTE SEDIMENTATION RATES
AND CELL VOLUMES IN MAMMALIAN BLOOD.**

**W. Drost-Hansen
Professor
Department of Chemistry**

**University of Miami
Coral Gables, Fl 33124**

**Final Report for:
Summer Research Program
Clinical Investigation Directorate
Wilford Hall Medical Center
Lackland AFB, San Antonio TX**

**Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, DC.**

September 1992

TEMPERATURE EFFECTS ON AQUEOUS POLYMER AND BIOPOLYMER SOLUTION VISCOSITIES, ERYTHROCYTE SEDIMENTATION RATES AND CELL VOLUMES IN MAMMALIAN BLOOD.

Professor W. Drost-Hansen

ABSTRACT

In spite of continued, extensive clinical use (especially overseas) , the dynamics and mechanism of the Erythrocyte Sedimentation Rate (ESR) remain poorly understood. As in previous years we have concentrated on elucidating some of the factors affecting the ESR through studies of the effects of temperature on the sedimentation process. In this connection we have extended the range of temperatures investigated to include far lower temperatures than previously considered: data are now available for many species down to about 8 °C. For the higher temperatures it appears that the dramatic changes with temperature seen near 45 °C in the ESR and Mean Cell Volumes (RBC) reflect changes in the structure of the interfacial water (the vicinal water) of the systems. For the ESR (and possibly for the MCV) we propose that the anomalous responses to temperature near 45 °C are most likely effected through the action of the vicinal water on the structural integrity of the spectrin molecule. Similarly, the effects of temperature on the rheological properties of aq. solutions of Polyvinyl pyrrolidone, Polyethylene oxide, Dextran, Bovine Serum Albumin (BSA), Fibrinogen and Cytochrome-c are intimately tied to the vicinal water of hydration of these polymers; thus, more or less dramatic changes in viscosity of such solutions are seen at (or very near) the "Drost-Hansen thermal transition temperatures," $[T_k]$, (for instance, near 15, 30 and 45 °C.) Similar transitions are seen in the viscosities of blood plasma from a number of mammalian species. In view of the ubiquitous nature of vicinal water it is hardly surprising that nearly all cell-physiological parameters - including osmotic and rheological aspects - vividly demonstrate thermal anomalies at T_k . Most likely, the involvement of vicinal water may explain many previously poorly understood aspects of the ESR while at the same time no doubt further complicating the kinetic picture of this process - for instance due to a distinct shear rate dependence [and hysteresis] of the nature and extent of the vicinal water. Finally, the thermodynamic properties of the intracellular water are also affected by the presence of the vicinal water; thus, the observed , anomalous cell volume changes with temperature (of both erythrocytes and platelets) reflect the influence of the vicinal water on the osmotic equilibria controlling cell volume regulation.

PLA. 007102

INTRODUCTION

After more than 70 years the Erythrocyte Sedimentation Rate continues to be used extensively in clinical and veterinary medicine (especially overseas.) However, the detailed dynamics of the ESR continues to escape a rigorous description although in principle many of the factors which influence the sedimentation process are well understood. The primary factors involved are the tendency for rouleaux formation (reflecting the RBC surface adsorption from the complex, macromolecular chemistry of the blood plasma), the plasma viscosity, shear rate, hematocrit, erythrocyte morphology, and a number of other factors of less direct influence.

We have continued our studies of the ESR over a wide temperature range; thus, while our previous measurements were restricted to approx. the range from 24 ° to 52 ° C, we have now extended the range down to about 8 °C for number of previously studied species (and a few new species.) At the same time, because of the expected Stokesian involvement of "particle size", we have continued our Mean Cell (RBC) Volume [MCV] measurements (and also Mean Platelet Volume [MPV] measurements) as functions of temperature. These volume measurements are highly useful for delineating the effects of temperature on the osmotic processes involved in volume control of the cell sizes. Finally, because of the importance of the plasma viscosity we have also measured viscosities of aqueous solutions of a variety of polymers: both model polymers (such as Polyvinyl pyrrolidone, Polyethylene oxide and Dextran) and some biochemically important macromolecules (such as Bovine Serum Albumin [BSA], Fibrinogen and Cytochrome-c.) Anomalous thermal responses continue to be observed at the Drost-Hansen thermal transition temperatures (T_k , near 15, 30, and 45 °C) in nearly all parameters measured!

PART I. ESR MEASUREMENTS

Erythrocyte sedimentation rates have been measured in whole blood from a number of healthy humans and various mammals, including some not investigated in our earlier studies. In some cases, the range of the ESR measurements have been extended to include lower temperatures, namely to encompass the 15 ° C anomaly of the vicinal water (i.e. the cell-associated water and the water of hydration of the aqueous macromolecules of the blood plasma.) As in previous years, the measurements are made using Wintrobe tubes placed in the Temperature Gradient Incubator (TGI = "polythermostat") for varying lengths of time. The results -- distance settled after X hours -- is plotted as a function of the Well number in the TGI. The Well # is

essentially proportional to the actual temperature (i.e. usually the temperature gradient in the bar is approximately linear.) In all graphs shown for ESR, MCV and MPV data, a few temperatures are indicated in the graphs; other temperatures may be estimated by linear interpolation. A typical example of a sed. curve is shown in Fig. 1; data for a killerwhale after incubation for 4.0 hrs.

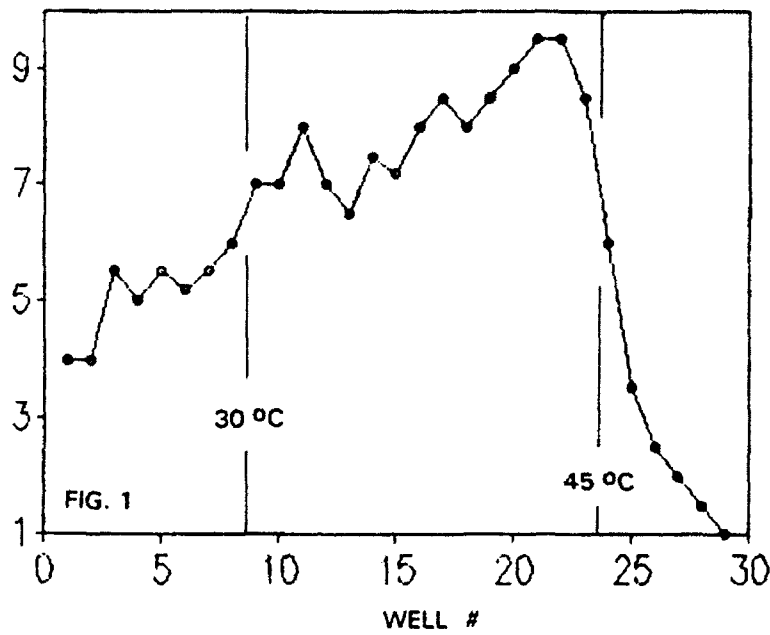
In general, the ESR curves closely resemble those obtained previously for the "high-temperature range" (about 24 - 54 °C.) Some variability has been observed in the case of the non-human samples: very high sed rates have been noted in samples from possibly sick animals as well as some cases of extremely low sed rates, — for instance for a dog (run 92TGI10B) with a sed rate of < 1 mm after 4 hours at 25 °C. However, in all cases studied a dramatic decrease in sed rate occurs for temperatures above 44 - 46 °C, regardless of the species, suggesting as before a distinct effect of the vicinal water. Blood plasma viscosities also change near 45 °C [please see Part V: viscosity data] but not nearly enough to account for the dramatic decrease in the sed rates. On the other hand, erythrocyte morphology does change at this critical temperature and we tentatively propose that the effects of the structural changes in the vicinal water is manifested via its influence on the red cell-associated spectrin. A considerable literature now exists which strongly suggests that spectrin plays a major role in stabilizing the red cell morphology. The spectrin molecule has a MW of about 240 000 Dalton (and may form dimers and possibly larger aggregates). Thus it is hardly surprising if the structural aspects of the vicinal water of hydration of the spectrin influence the stability of the spectrin. At the thermal transition temperature, the vicinal water is believed to undergo a structural phase transition from one stabilized structure below 45 ° to another, stable but different structure above this critical temperature. Most likely the transition state between these two stabilized forms is characterized by notably increased disorder, thus likely resulting in a concomitant decrease in the stability of the spectrin conformation/association. This destabilization of the spectrin "framework" may thus lead to morphological changes of the RBC -- vividly visualized in photomicrographs of the RBC after exposure to the higher temperatures [see below.]. The tendency to rouleaux formation must be significantly reduced due to the morphological changes and this in turn is likely the primary reason for the drastic decrease in the sed rate above the critical temperature. Furthermore, changes are also expected in the nature and extent of protein adsorption on the surface of the RBC, again affecting the degree of rouleaux formation and thus affecting the sed rate.

Measurements of sed rates around the two lower vicinal transition temperatures, 15 ° and 30 ° C, sometime reveal changes at these critical temperatures but these effects are far smaller than for the 45 ° anomaly, and in fact, sometimes not seen at all. Two relatively typical examples of sed rates at low temperatures are shown in Fig. 2 and Fig.3 - respectively for a dog and a Yucatan pig. The role of the vicinal water in these cases may be primarily via the influence on the vicinal hydration of the plasma proteins and the cell membranes. This suggestion is consistent with the frequently observed anomalies at T_K seen in the viscosity measurements on various protein solutions and blood plasmas. Likewise, the vicinal water may also influence the extent of erythrocyte surface adsorption of the proteins at low temperatures involved in the rouleaux formation.

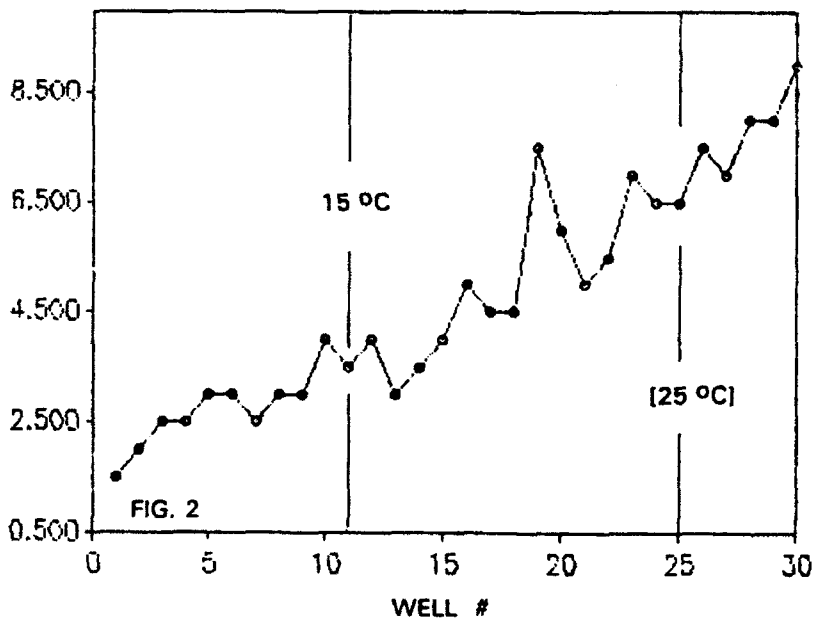
Less pronounced suggestions of anomalies are also sometimes seen in sed rates near 21 (+/- 1) °C. Such sed rate anomalies have previously been reported also by Prof. R. Glaser et al. (of Humbolt University, Berlin) and ascribed by these authors to membrane bound lipid phase transitions.

Finally, in some - but by no means all - sed rate measurements, indications of anomalies may also been seen near 37 ° C. Again it is tempting to suggest that these anomalies, like the one near 21 °, may be the result of a (membrane) lipid phase transition. On the other hand it is difficult to envision why anomalies should occur near this temperature as 37 ° - 38 ° C is close to the body temperature of most mammals. It would seem impractical during evolution to allow a homeostatic system to operate near a critical temperature where the changes in properties over a narrow temperature range might lead to notably different stabilities. In the future this point deserves close attention because of the possible clinical implications of body temperature control in the normal and febrile patient. Because of the excellent reproducibility and precision of our sed rate measurements and Mean Cell Volume data such measurements may prove uniquely well suited for a detailed study of the blood around body temperature. Thus it would be easy to program the Temperature Gradient Incubator to allow (nearly) simultaneous measurements from, say, 32 ° C to about 42 ° C, thus providing great resolution with data points spaced only 0.3 ° apart.

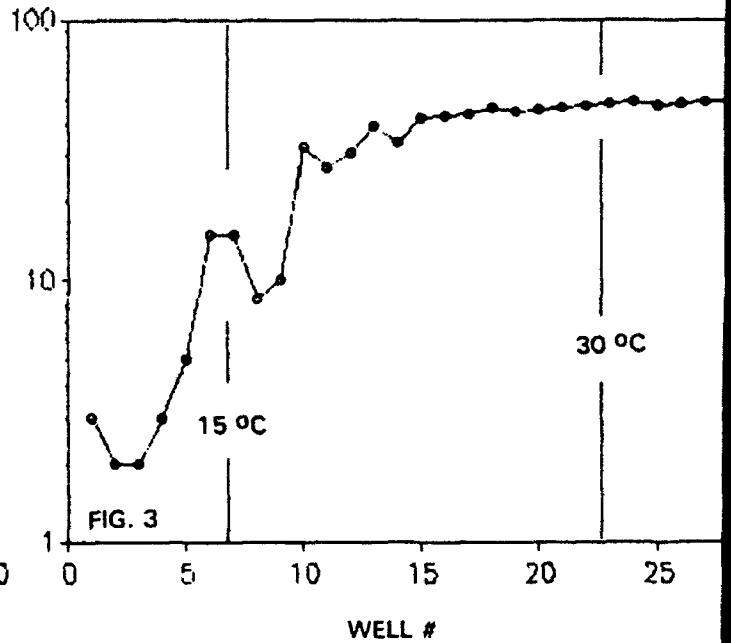
DISTANCE SETTLED (mm); FOR $\Delta t = 4.0$ hr.
KILLERWHALE (MALE). 92TG119B



DISTANCE SETTLED (mm); FOR $\Delta t = 4.0$ hrs
DOG 92TG104B



LOG [DISTANCE SETTLED, mm]; FOR $\Delta t = 1.0$ hr.
YUCATAN FIG. 92TG108B



PART II. MEAN (RED) CELL VOLUME MEASUREMENTS.

Mean Cell Volume (MCV) measurements have been continued this year both because of the potential role of the cell volumes in determining the ESR and because of the intrinsic interest of cell volumes for an understanding of the osmotic processes in living organisms. As before, the MCV data were obtained with a Baker, Model 9000, Hematology Counter. Usually, MCV measurements were made immediately after the whole (anticoagulated) blood was removed from the Temperature Gradient Incubator (the TGI) after incubation at the various temperatures. From a series of separate measurements it was shown that -- regardless of the range of incubation temperatures, - the MCV (and MPV as well) remained unchanged for periods of at least two hours after removal of the samples from the TGI with storage at room temperature. As in earlier years, the reproducibility of the cell volume measurements was excellent, usually better [or far better] than 1 %; this is amply illustrated by an inspection of the rather smooth curves of cell volumes as a function of temperature.

As observed in our previous studies, the MCVs invariably show dramatic changes at 44 - 46 °C! Obviously something dramatic happens in this temperature range and the degree of abruptness suggests that a phase transition is most likely responsible. As there are no major lipids with a phase transition at this temperature it is almost inescapable that the transition is due to the vicinal water with its well-documented transition (T_k) at 45 °C. However, in the past it was not at all obvious, on a molecular scale, where is the "site of action" of the vicinal water. As discussed above (under ESR) it now appears likely that the red cell structural protein spectrin is the moiety most strongly affected by the structural transition of the vicinal water. [We wish to thank Professor J. Stuart of University of Birmingham, UK, for this suggestion.] If indeed the configuration of the spectrin molecule is notably destabilized at the temperature where the vicinal hydration structures change it would be reasonable to expect that the structural integrity of the red cell membrane suffers.

Typical examples of the effects of temperature on MCV are shown in Fig.4, Fig. 5 and Fig. 6 from this years series of measurements. Fig.4 shows the data for MCV from a goat; note the rather small erythrocyte volumes. Below 46 to 47 °C, the observed volumes are nearly independent of temperature: $17.85 \pm 0.1 \mu^3$ while the volumes increase dramatically above 47 - 48 °C. Similarly, in Fig 5 , for the case of blood from a seal -- nearly complete temperature independence of MCV below 45 °C with an abrupt increase above this temperature (followed by a very distinct decrease) Finally, shown in Fig. 6 are the results for blood from a killerwhale, Note the large erythrocyte volumes but again no temperature variations below 44 - 45 °C, followed by the same type of pronounced MCV increase above this temperature and a subsequent

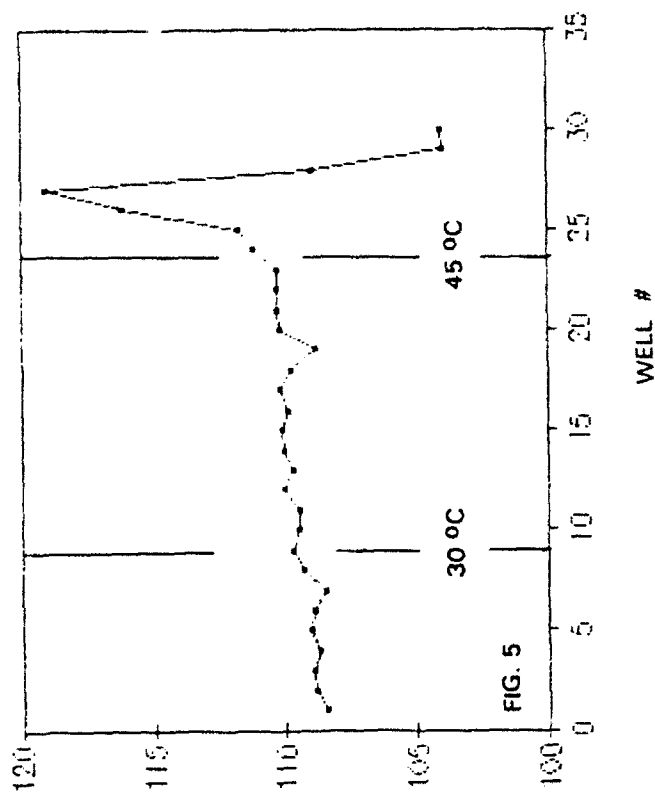
drop above 49 °C. Our data on humans generally show the same behaviour as for seals, and killerwhale..

In the lower temperature range which we have now investigated it is found the MCV is generally quite unaffected by temperature. For blood from pigs the variation in MCV over the range of 8 to 35 °C is within ± 0.2 % (!) [with no discernable trend.] Somewhat similar, Fig. 7, and Fig. 8 show MCV data for blood from, respectively, a dog and a Yucatan Pig. An inspection of these and other data obtained this summer suggests some systematic trends:: for human subjects an apparent increase in MCV above approx. 30 °C although this is not always seen. The MCV data for the Yucatan Pig decrease slightly with increasing temperature while the data for the dog show a distinct, if modest, increase up to about 25 °C followed by a slight decrease. In most of the above cases the data suggest that some real trends may exist but overall, for the cases discussed here, the net change with temperature is modest, less than 2 % - and in some cases much less.

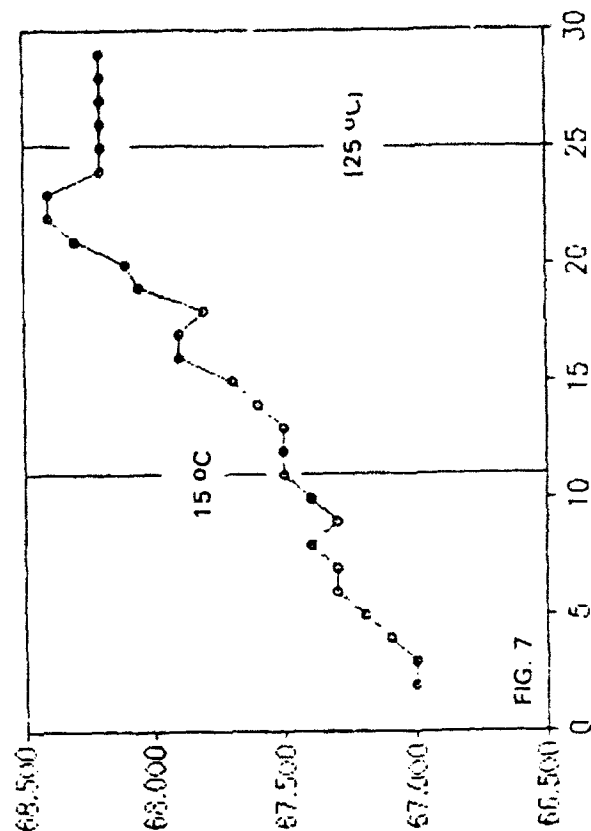
As mentioned earlier in this report, reasons exist to suspect the occurrence of anomalies at or near 15 and 30 °C, related to vicinal water structure changes, while an anomaly may exist near 21 °C because of a membrane lipid phase transition at this temperature. Unfortunately, the total MCV changes seen here are generally too small, compared to the precision of the data, to allow an interpretation of the current data in terms of a structural membrane transition at (or very near) 21 °C. We hope to be able to pursue the question of all the low temperature thermal anomalies (related to vicinal water and membrane lipids) in cell volumes at a later time.

FILE: COPIES

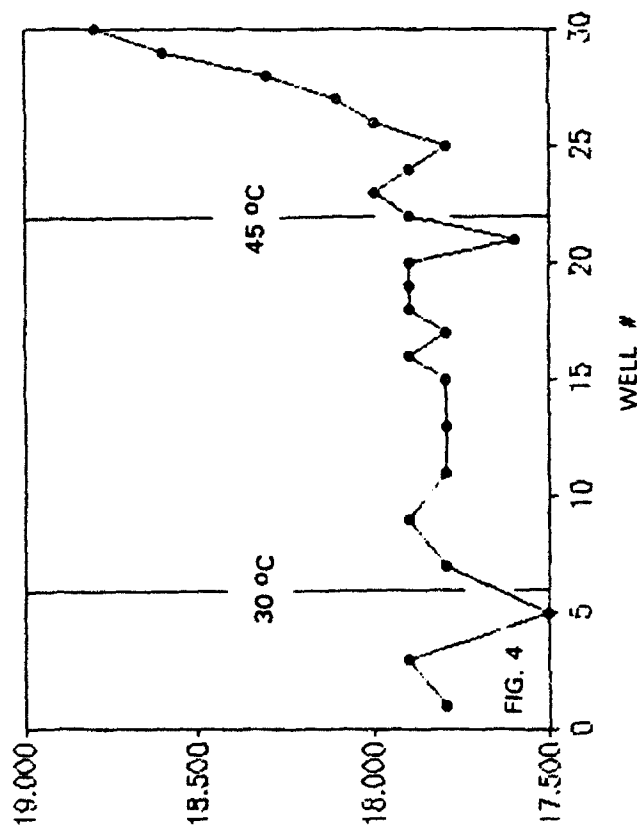
MCV (μ^3) SEAL 92TGI22B (# 2, FEMALE)



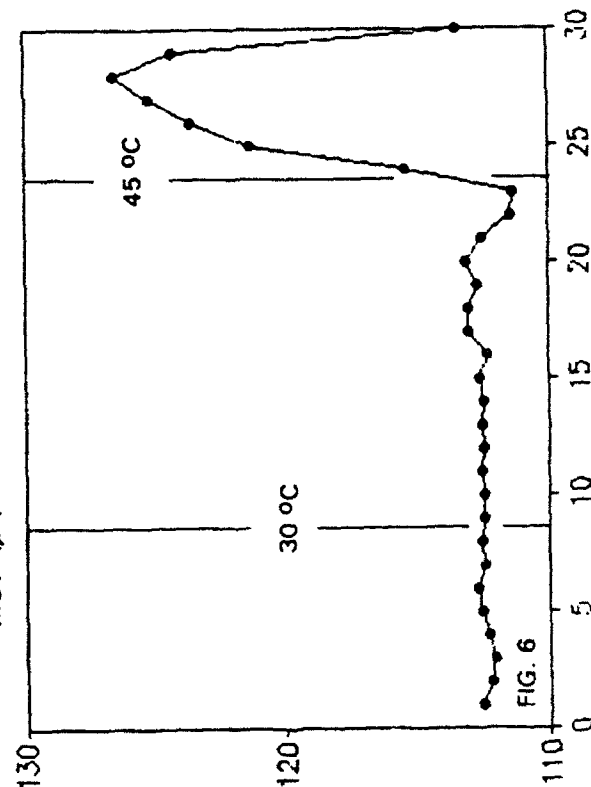
MCV (μ^3) DOG 92TGI04B



MCV (μ^3) GOAT 92TGI03B



MCV (μ^3) KILLERWHALE (MALE) 92TGI19B



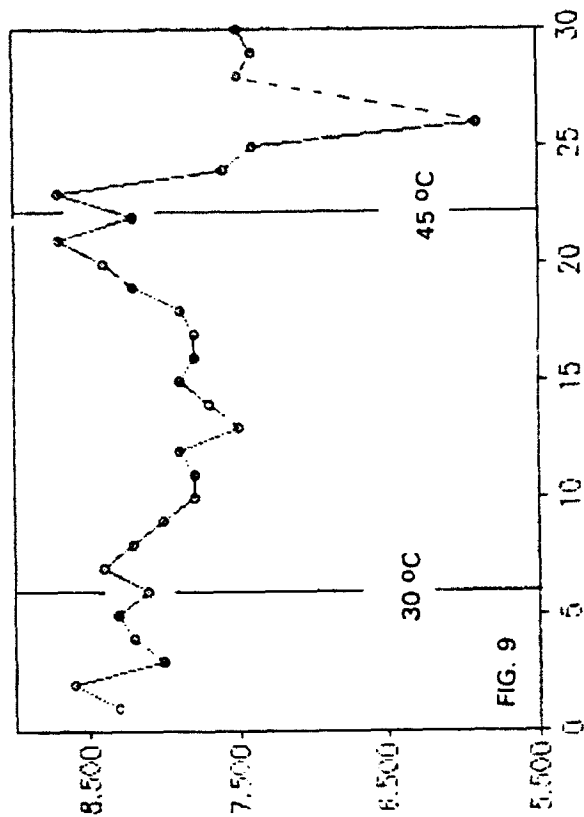
PART III MEAN PLATELET VOLUME MEASUREMENTS

Mean Platelet Volume (MPV) measurements have been made this summer in connection with all the TGI runs (92TGI01B through 92TGI22B). The effects of temperature on MPV previously observed at high temperatures (say, 43 to 55 °C) have been confirmed in the experiments carried out this year. Regardless of the species studied, the most pronounced changes are usually abrupt increases in the MPV around 44 - 48 °C followed by sharp decreases in the volumes. Part of these volume changes likely signal osmotic effects on the platelets while part of the changes may reflect the observed, strong tendency for "clumping" of the platelets above the critical temperature range of 44 - 46 °C. [please see discussion of blood cell morphology, Part IV.] Fig.9, Fig. 10 and Fig. 11 show typical sets of data for MPV, respectively for a human, a goat and a killerwhale.

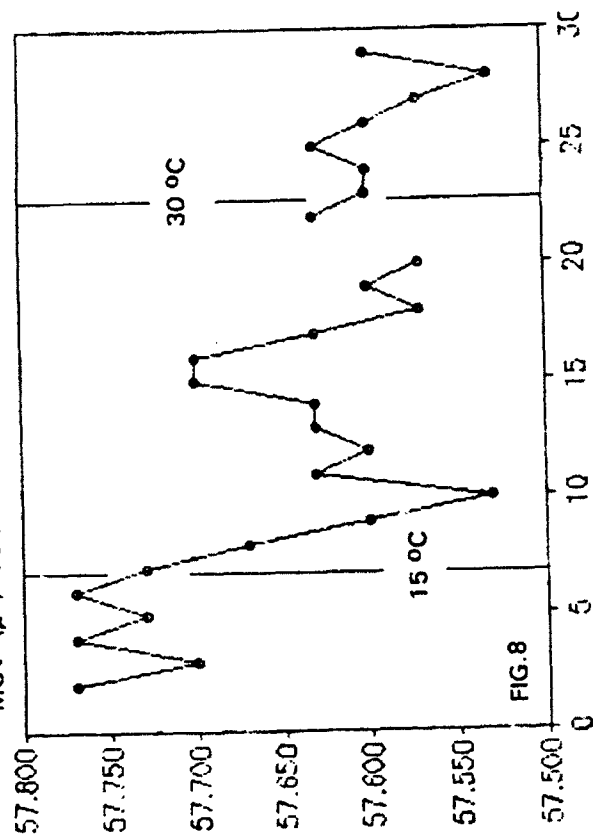
A number of low temperature measurements of MPV have also been made over the range from about 8 to 35 °C. Typical changes in MPV are about 10 % over this temperature range which is notably larger than the volume changes observed over the same temperature range for erythrocytes (generally less than 1 (to 2) %.) Two typical sets of data are shown in Fig.12, Fig. 13 and Fig. 14, respectively, for a dog, a Yucatan pig and a Rhesus monkey.

PLA C001R10

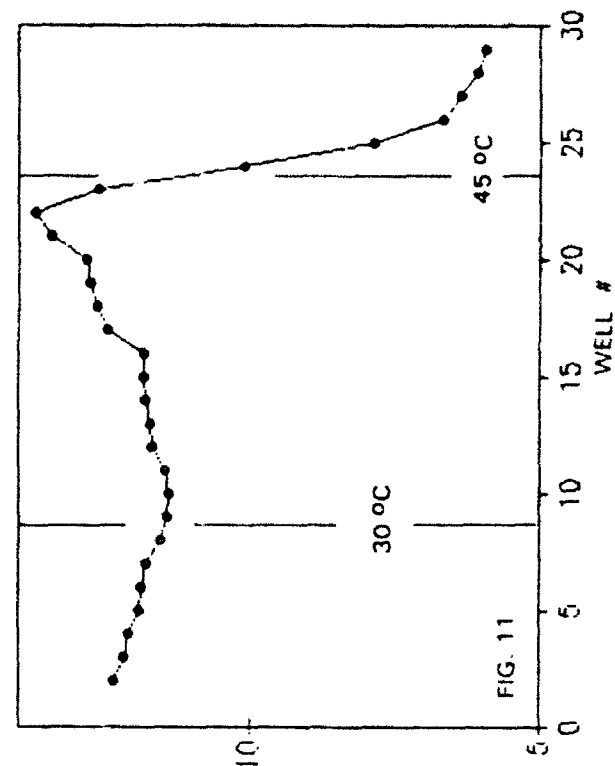
MPV (μ^3) SPECIMEN: HUMAN (D-H) 92TG101B



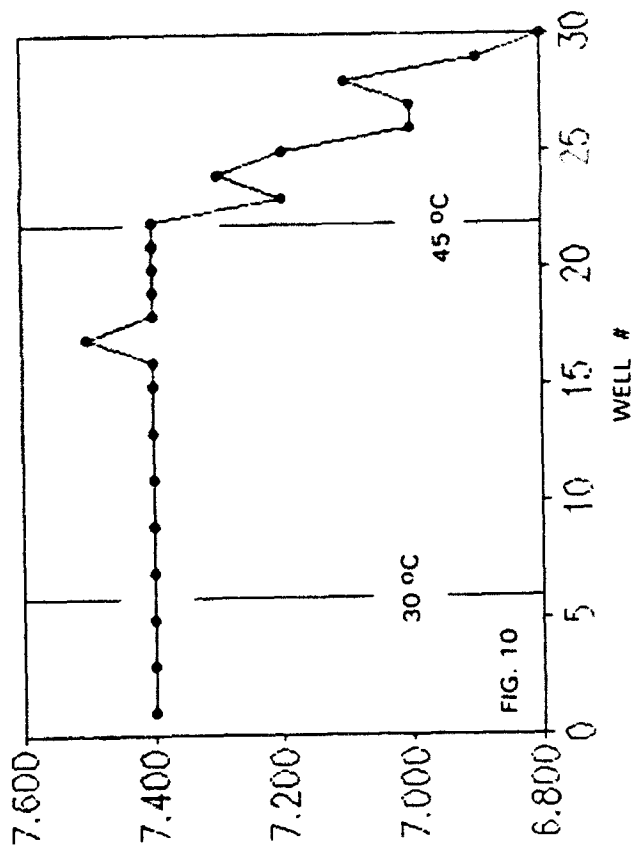
MCV (μ^3) YUCATAN PIG 92TG108B

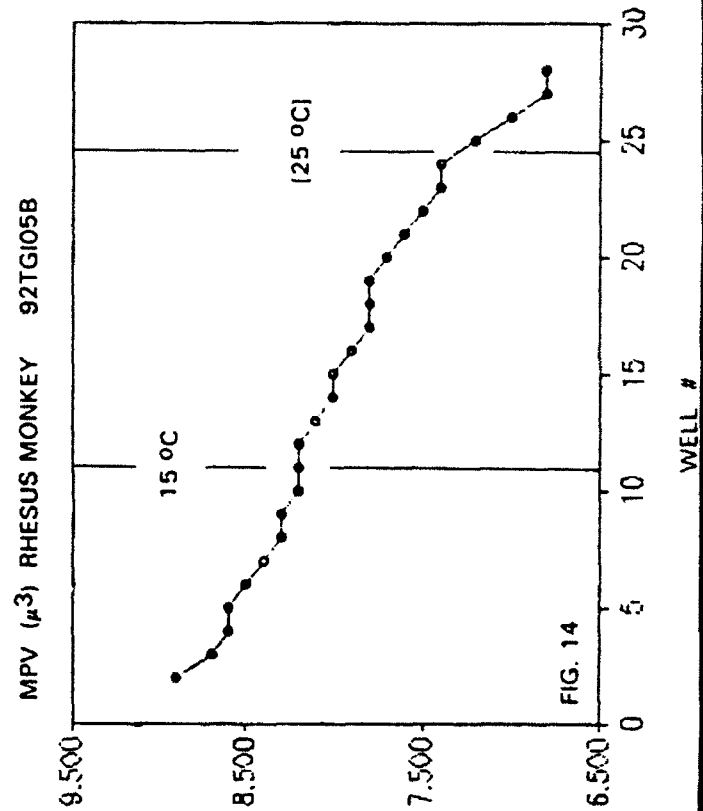
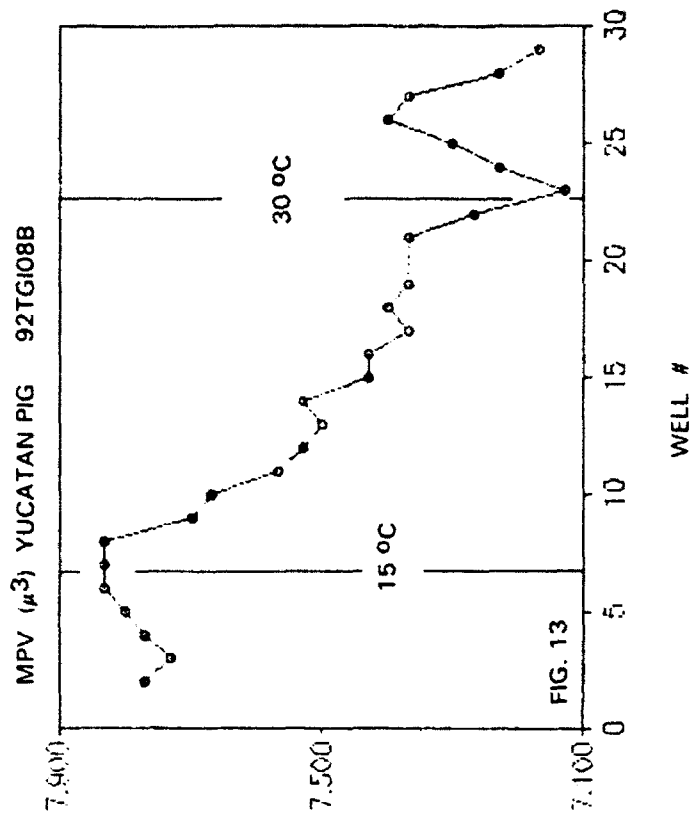
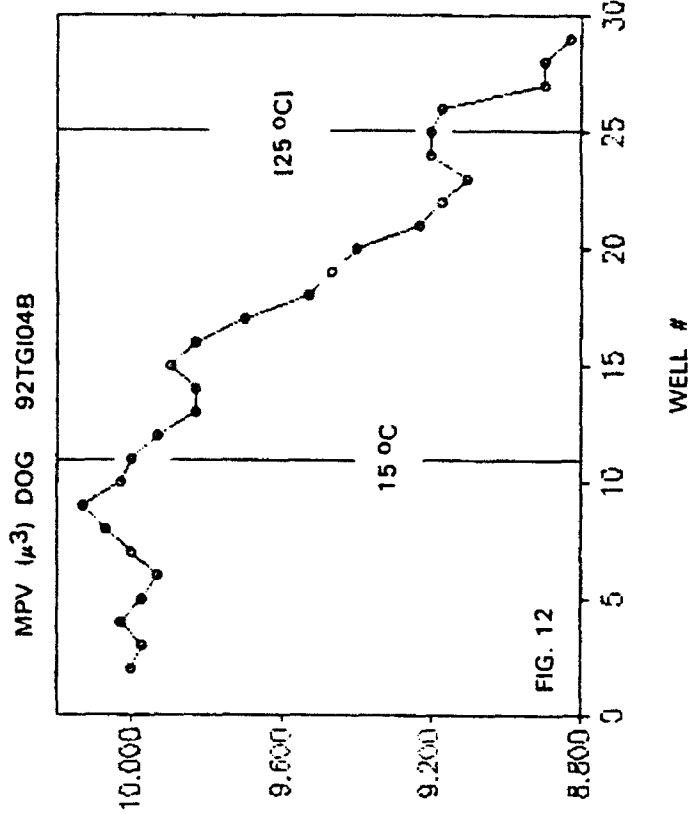


MPV (μ^3) KILLERWHALE (MALE) 92TG119B



MPV (μ^3) GOAT 92TG103B





PART IV: BLOOD CELL MORPHOLOGY

In previous studies (1991) notable changes in morphology were observed in a variety of blood samples exposed to temperatures of 44 °C and above, consistent with the findings in a number of reports in the literature. This summer, 1992, we have carried out a more detailed microscopic study of heated blood samples; in particular, careful observations have been made on blood cell morphology in two experiments on human samples from healthy individuals, respectively from a young male and a female; [Runs 92TGI17B and 92TGI18B]. Blood samples for microscopic examination were taken after two hours of incubation at the various temperatures. The microscopy and analysis of the findings were carried out by a Hematologist, Lt.Col. Wayne Patterson [Ph.D.]. The findings from the two set of samples were nearly identical; below are presented a synopsis of the observations made on sample 92TGI18B.

25° to 42 °C: All cells appear normal in size and shape. Platelets are singular and show granularity consistent with a non-activated state. Neutrophils display a normally appearing segmented nucleus and fine, even granularity.

43.4 °C: RBC's appear to be "stacking" somewhat and some are displaying a shape change to spherocytes and helmet shaped cells. Platelets still appear essentially normal but with some evidence that clumping is beginning to happen. Neutrophils have essentially normal segmentation but slight toxic granulation present in the cytoplasm.

44.4 °C: RBC stacking and tendency towards spherocytosis are seen. Platelets maintain granularity but beginning to aggregate. The neutrophil normally segmented but the cytoplasm is becoming more basophilic.

45.4 °C: RBC stacking evident; platelets are clumping. The neutrophil shows more basophilic cytoplasm and moderate toxic granulation.

46.4 °C: RBC shape becoming more anomalous. Helmet cells more prevalent. Platelet aggregation is evident and at least half of the platelets are granular, indicating activation and secretion. The neutrophil nucleus is more pycnotic and the cytoplasm more basophilic (which usually indicates cell death and degeneration.)

47.4 °C: RBC shape is even more anomalous with helmet cells, spherocytes and acanthocytes present. "Blebs" on several of the RBC indicate membrane reorganization, probably due to protein conformational changes (denaturation.) Platelets are mostly agranular and aggregated. The neutrophil nucleus

is quite pycnotic and the cytoplasm is very basophilic, displaying moderate toxic granulation. Some platelets adhere to neutrophils.

48.4 °C: Normal RBC shape is practically non-existent. Helmet cells, spherocytes, RBC membrane fragments and other bizarre shapes are present, almost exclusively. Cytoplasmic blebbing is widespread. Platelets are aggregated and mostly agranular.. Both the nuclear and cytoplasmic membrane of the neutrophil have disintegrated; the cell itself is likewise disintegrating.

49.4 ° - 51.4 °C: All findings resemble the observations made at 48.4 °C.

The microscopic findings again suggest that dramatic events on the molecular scale occur near 45 ° to 47 °C. In all probability the major event is the 45° transition of vicinal water which most likely corresponds to some kind of transition from one stable structure below the critical transition temperature to another, stable structure above this transition range. It is surmized that a strongly disordered state of the vicinal water prevails right at the transition region, say within a temperature interval of one (to three) °C. As implied in the earlier discussion in this report, the structural membrane protein spectrin may be the major element to become destabilized at T_k . The morphological changes observed go a long way towards explaining the observed sed. rate anomalies and the observed changes in MCV and MPV.

FILE: G010111

PART V. VISCOSITY DATA

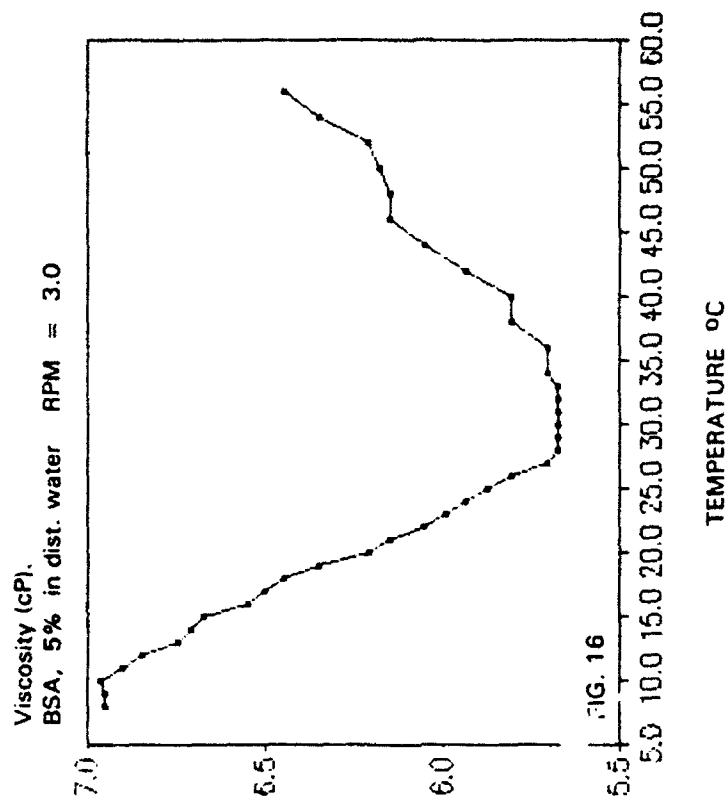
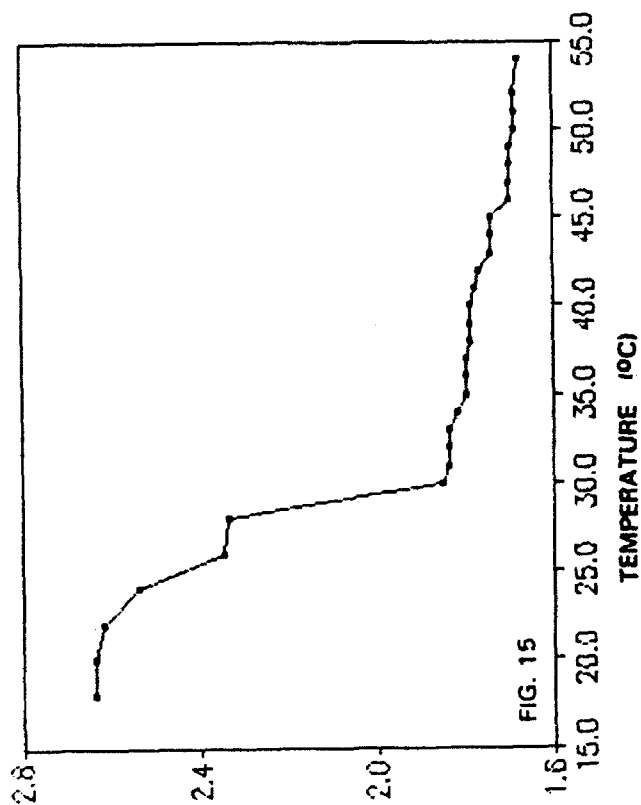
Viscosity measurements, using a Brookfield Model LVTDV-22 variable shear rate viscometer, have been made on aqueous solutions of the following polymers: Polyvinyl pyrrolidone, Polyethylene oxide (PEO), Dextran, Bovine Serum Albumin and Cytochrome-c; - as well as measurements on blood plasma from a number of mammalian species.

The viscosities of most of these systems are sensitive to one or more of the following variables: concentration, ionic strength, pH, shear rate, temperature, past thermal history, age of the solutions and possibly several other parameters not readily identified or controlled. Typical results are presented in Fig.15 through Fig. 18. The one persistent feature in all of these graphs (and indeed in ALL data sets obtained on aq. polymer solutions this summer and in 1991) is the occurrence of thermal anomalies at (or quite near) the Drost-Hansen thermal transition temperatures, T_k , of vicinal water. Given the notable diversity of the polymers studied, it appears that the only element these macromolecules are likely to have in common is the solvent -- or, more specifically, the water of hydration. Thus it is proposed that consistent with our previous prediction (Etzler and Drost-Hansen, 1983) ALL macromolecules in aq. solution are likely vicinally hydrated and exhibit the thermal transitions of the vicinal water. (Note that the same type of thermal anomalies were also observed in our 1991 AFOSR study of aq. fibrinogen solutions.)

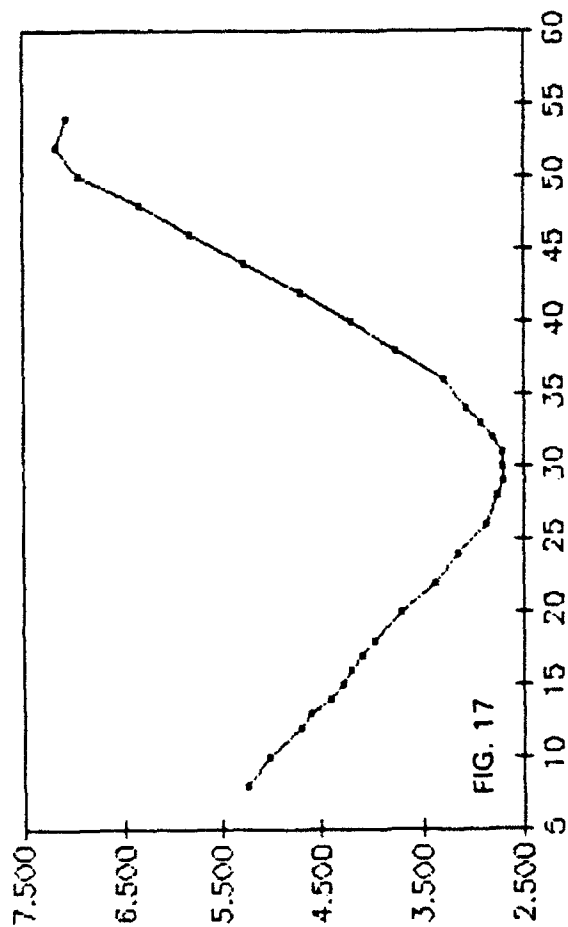
We have previously demonstrated that most (or indeed likely ALL) solid interfaces induce vicinal water structures and as living cells provide a vast amount of intracellular interfaces they must as a result contain a great deal of vicinal water. Add to this the conclusion reached above: that all large macromolecules in aqueous solution are also vicinally hydrated, and it must be concluded that the effects of vicinal water must ramify through ALL biological cells and that many of the thermal anomalies reported above in Erythrocyte Sedimentation Rates, Mean Cell Volumes and Mean Platelet Volumes are indeed manifestations of the cell-associated, vicinal water.

FILE: CDP1988

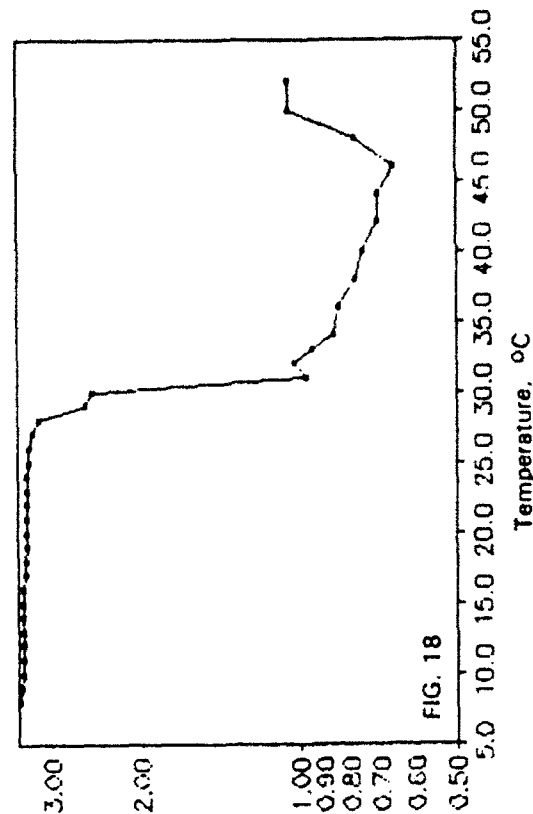
log [viscosity (cP)]
DEXTRAN, 5 %, RPM = 12, 6-11-92



Viscosity (cP).
BSA, 5% in 0.9 % NaCl. RPM = 3.0



log [viscosity(cP)] vs. temp.
PLASMA Harbor Seal 6.0 RPM, 7-10-92



CONCLUSIONS

The Erythrocyte Sedimentation Rate in whole blood has been found to drop abruptly for temperatures above 45 - 46 °C for blood from all species studied, consistent with our findings from 1991 and 1989. This dramatic effect is believed related to the change in vicinal water structure at that temperature (one of the "Drost-Hansen transition temperatures", T_k .) The transitions in vicinal water near 15 and 30 °C do not appear to affect sed. rates to any great extent although some evidence for anomalies in the sed. rates at these temperatures have been seen.

As in previous years, remarkable changes are found in Mean Cell (Red Cell) Volumes (MCV) near 45 - 46 °C. In some cases, notably less dramatic changes are seen also near 15 and 30 °C but in general MCV values are remarkably unaffected by temperature (typical changes less than 1 (to 2) % over a 30 degree interval, below 35 °C.)

Measurements of Mean Platelet Volumes (MPV) invariably show large variations with temperature near 45 - 46 °C. At lower temperatures (say less than 35 °C) MPV values change far more with temperature than observed with Erythrocytes (almost an order of magnitude more) but only occasionally do the thermal anomalies near 15 and 30 °C seem to affect the platelet volumes.

Blood cell morphologies are dramatically affected by temperatures around 44 - 48 °C: Erythrocytes become grossly distorted; neutrophils disintegrate and platelets aggregate. It is almost certain that these changes must reflect the thermal transition in vicinal water at that temperature range. It is proposed that one of the major red cell structural proteins, spectrin, may be particularly sensitive to the structural aspects of the vicinal water.

As observed last year (1991) we have also this summer found anomalies in the viscosity of aqueous polymer solutions near the thermal transition temperatures of vicinal water, T_k , i.e. near 15, 30, 45 and 60 °C. In any given run, only one or two of the thermal anomalies may be seen in the viscosity data [although in some runs all the anomalies are seen] However, in NO case have we ever failed to see at least one of the vicinal water transitions (1991, 1992.)

Because of the diversity of the polymers studied (: PEO, PVP, Dextran, BSA, Fibrinogen, Cytochrome-c and blood plasma and serum) the thermal anomalies seen are unlikely to reflect some polymer-specific aspect. Instead, it is reasonable to propose that the anomalies must be caused by the one element in common for all these solutions, namely water. As bulk water definitely does not show thermal anomalies the only possible other element must be the vicinal water of hydration of the polymers. In view of this and the overwhelming evidence for vicinal water at all solid/water interfaces, it is inescapable that vicinal water must ramify through ALL of cell biology (and medicine.)

REMMONDATIONS

Let us do more ! -- especially explore the role of vicinal water in clinical medicine.

SPECIES STUDIED AT CID: 1989, 1991, 1992.

ESR, MCV, MPV

as well as viscosity measurements

on some model polyelectrolytes, biopolymers, blood plasma
and serum and Na/K distributions as functions of temperature.

For most of the species listed below, ESR, MCV and MPV data have been obtained (with the exception of ESR data for those rather rare cases where the sed rates were too low to yield significant information..) In nearly all cases studied, data were obtained over the range of approximately 22 to 52 °C and during the 1992 period a number of measurements were also made between 4 and 35 °C.

HUMANS

CHIMPANZEE
BABOON
MONKEY
PIG
YUCATAN PIG
DOG
HORSE
CAT
COW
SHEEP
GOAT
LLAMA
RAT
RABBIT
KILLER WHALE
HARBOR SEAL

As of the time of preparation of this report (Sept. 1992) five manuscripts are in preparation for publication. These papers will deal, respectively, with: A) ESR results; B) MCV data; C) MPV data; D) viscosity data; and E) time-of-storage effects on sed rates and cell volumes in whole blood. In addition, a major presentation, describing vicinal water and the hydration of polymers in aqueous solution, was made at the Gordon Research Conference on Water, Aug. 1992.

FILE: CID9901

ACKNOWLEDGEMENT

The author wishes to thank his colleagues and coworkers, Col. J.H. Cissik, Ph.D., and Lt. Col. W.R. Patterson, Ph.D., for their wonderful, continued help, advise and encouragement. The author also wishes to express his sincerest thanks to the following persons for their important, logistics support in a wide variety of fields: Dr. William Ehler, Mr. David McGlasson, Ms. Rita Smith, Mr. David Boucher, Ms. Cathryn Smith, Ms. Patti Allen, Mr. Glenn Kettering, Mr. David Goddu, Mr. Michael Chastain, Ms. Maureen Lorenzen, Ms. Cathy Smith, Mr. David Boucher, Mr. Bennie Farrel MSgt. James Joseph and his administrative staff; and particularly the Chief of the Biological Sciences Division of CID, Maj. Donald Bradley, (Ph.D.). Thanks are also due to Dr. Les Dalton of SeaWorld of Texas, San Antonio, TX., for providing blood samples from various aquatic mammals.